AF IFW

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	)	
James H. Prestegard	)	Art Unit: **1631**
	)	
Serial No. **09/880,648**	)	Examiner: **Lori A. CLOW**
	)	
Filed: **June 13, 2001**	)	
	)	
For: **NMR Assisted Design of High Affinity Ligands for Structurally Uncharacterized Proteins**	)	

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

APPELLANT'S BRIEF UNDER 37 C.F.R. § 1.192

CORRECTED

In support of the appeal entered in the above referenced application, and in response to the Notice of Non-Compliant Appeal Brief mailed on October 21, 2005, Appellant hereby submits this brief under 37 C.F.R. § 1.192.

Real Party in Interest

The real party in interest is the inventor – Dr. James H. Prestegard.

Related Appeals and Interferences.

None.

Status of Claims

Claims 1-4, 6-16 and 19-20 stand finally rejected. Appeal is taken from the rejection of

---

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Mail Stop AF, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450, on December 20, 2005.

Clark G. Sullivan, Reg. No. 36,942

all claims.

Status of Amendments

Amendments made in Applicant's First Amendment After Final, mailed on December 27, 2004, have been entered for purposes of appeal, and are reflected in the Claims Appendix attached hereto.

Summary of Claimed Subject Matter

The invention is a new process for constructing biologically active ligands with improved specificity and binding affinity for biological targets, by starting with a ligand that is known to bind a biological target, identifying a second ligand that binds to the biological target, and constructing a composite ligand from the two ligands in which the distance between the two ligands in the composite ligand, and the orientation of the ligands to one another in the composite ligand, closely mirrors their relative distance and orientation when bound to the substrate as separate ligands.

In a principal embodiment, described in the specification as the "Direct Method" on pages 9:24-10:14, the process is carried out according to the following steps:

1. paramagnetically labeling a ligand that is known to bind to a biological target;
2. preparing a complex between the paramagnetically labeled ligand and the biological target;
3. preparing NMR spectra of the complex;
4. modifying the complex by the addition of a second ligand that binds to the biological target;
5. preparing NMR spectra of the modified complex; and
6. analyzing the spectra to determine whether the second ligand has bound to the biological target within the paramagnetic zone of the label on the first ligand;
7. further analyzing the spectra to determine the distance of separation between the first and second ligand when bound to the biological target;
8. deducing the relative orientation of the first and second ligands when bound to the biological target; and
9. linking the two ligands together at the distance of separation and orientation deduced

in steps 8 and 9.

The combined ligands have a greater specificity and affinity for the biological target than either of the ligands by itself.

Independent Claim 1 -- Independent claim 1 covers the "Indirect Method" described above and in the specification at 9:24-10:14, but broadens it slightly to cover the several ways that the NMR spectra can be generated for the biological target and the second ligand. For example, the spectra can be generated from a complex formed between the biological target and the second ligand, and compared to the spectra from the complex formed between the biological target and the first ligand, or spectra can be generated from a tri-part complex formed between the biological target, the second ligand and the first ligand, and compared to the spectra from the complex formed between the biological target and the first ligand. Support for this expansion is contained in the specification at 10:26-10:30.

Independent Claim 19 -- Independent claim 19 covers a variation of the foregoing process, referred to in the specification as the "Direct Method," in which NMR spectra are also prepared of the second ligand and a paramagnetically labeled derivative of the first ligand in the absence of the biological target, and compared to spectra derived from a tri-part complex formed between the biological target, the second ligand and the first ligand, and spectra derived from a complex of the biological target and only one of the ligands, to deduce the relative distances of the ligands when bound to the biological target. This method is described in the specification at 11:1-11:18.

Grounds of Rejection to be Reviewed on Appeal

The following issues are presented on appeal:

(1) whether the preamble in claims 1 and 19 is unclear because it recites "a method for improving the binding affinity of ligand for a biological target," when the claim does not specifically recite an improvement in the binding affinity.

(2) whether step (c) of claim 20 is rendered unclear by the presence of the term "substantially."

(3) whether the claims are properly rejected as obvious over the combined teachings of Johnson et al. (1999) and Bolon et al. (1999).

Argument

Claim Rejections – 35 U.S.C. § 112, second paragraph

The Final Office Action rejects claims 1 and 19 because they recite a method for improving the binding affinity of a ligand for a target and do not provide a step of improving the binding affinity. This rejection is improper because (1) the objectionable limitation is found in the preamble, which does not even limit the claim, and (2) the binding affinity is inherently improved in a significant number of constructs whenever one carries out the recited sequence of steps.

The Final Office Action rejects claim 20 because clause (c) contains the term "substantially." Applicant respectfully submits that a skilled worker would understand precisely what is meant by the term substantially, since it refers to a distance and orientation that improves the binding affinity of the resulting compound.

Claim Rejections – 35 U.S.C. § 103

The Final Office Action also rejects the claims as obvious based on the combined teachings of Johnson (1999) and Bolon (1999). During a telephone interview conducted on March 15, 2005, the Examiner informed Applicant's attorney that she was relying on page 620, in column 2 of Johnson, which describes composite protein that Johnson constructed by linking together two cellulose binding domains (CBD_{N1} and CBD_{N2}) of the CenC beta-1,4-glucanase enzyme. This linking of two protein domains is in contrast to the linking of two parts of a ligand that binds to a protein as described in the instant claims. Johnson was studying the composite protein with the two CBDs to gain some understanding of the binding characteristics of the CBDs to cellulose, but ultimately he concluded that the relative orientation of the CBD protein domains did not even matter, because "the tandem CBDs anchor CenC to its natural substrate … without a strong preference for their orientations." (P. 620 at col. 2.)

Johnson did not link the CBD domains of the composite protein at a preferred orientation and distance to improve their binding affinity for a ligand, or undertake any steps to characterize that preferred distance and orientation, all as required by the claims of the present invention. He does not identify the structure of the linker, the distance between the two CBDs when bound to the linker, or the relative orientation of the two CBDs when bound to the linker. In fact, he states just the opposite in column 2 of page 620 when he states that "the spatial arrangement of these

modules (i.e. the two CBD ligands) within native C. *fimi* CenC remains to be defined." Continuing in the second column of page 620, Johnson further confirms the point when he states that the two ligands "cannot bind simultaneously to adjacent regions of a single polymer chain" due to "structural constraints" in the linker.

Johnson never addressed the issue of a distance between a first and second ligand and never "determine[ed] whether a second ligand perturbs peaks on the second NMR spectra that are also perturbed by the paramagnetic label on the first NMR spectra," as required by clause (d) of claim 1. That step would only be performed if one was interested in determining whether the two ligands bound to the target molecule at an ascertainable distance, and constructing a composite ligand at the distance ascertained. Moreover, no orientational data from dipolar couplings were collected as described in claims i.e. 7, 19(e) and 20(b). The "orientation" of the cellulose-derived ligand recited by Johnson actually refers to a "direction" of binding in the protein site which proves to be ambiguous, as opposed to a three dimensional orientation. This latter information was also dependent on use of assigned protein resonances, a step not needed in the present invention.

To overcome these deficiencies, the Examiner makes reference to the fact that Johnson reports some spin label studies that he conducted in column 1 of page 620. These studies were performed to locate the binding site of a single ligand on the protein surface, not to define distances between two ligands; again it required assignment of protein resonances. Once again, this represents a fundamental difference between the method Johnson was employing and the method described in the instant claims.

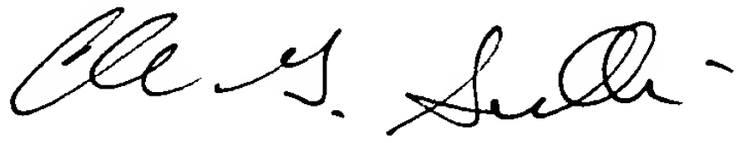
The Examiner also cites to Bolon (1999) to cure these deficiencies, but Bolon merely reports the availability of residual dipolar couplings to deduce the relative orientation of a ligand when bound to a biological target based upon observations of residual dipolar couplings. The article only analyzed the orientation of a **single** ligand relative to a protein structure. The authors did not bind two different ligands to the protein structure -- as required by the pending claims -- , deduce the relative orientations of the two ligands – as required by the pending claims -- or determine the distance between the ligands – again as required by the pending claims, because he was not concerned with improving the binding affinity of a ligand for the biological target by making composite ligands. In short, the reference did not offer any suggestion or motivation for constructing composite ligands based upon their relative distances and orientations when bound

to a protein, as recited in the pending claim. Lacking such a suggestion or other motivation, the references do not support a prima facie case of obviousness.

CONCLUSION

In view of the arguments presented herein, Applicants respectfully request that the final rejection in this matter be vacated, and that this application be returned to the examiner with instructions to enter a notice of allowance.

Respectfully submitted,



Clark G. Sullivan
Reg. No. 36,942

December 20, 2005
KING & SPALDING LLP
191 Peachtree Street
Atlanta, GA 30303
(404) 572-4600 (Telephone)
(404) 572-5145 (Facsimile)
K&S Docket No. 04342.105053

APPENDIX 1

CLAIMS APPENDIX

1) A method for improving the binding affinity of a ligand for a biological target comprising:
   a) preparing first NMR spectra of a first complex comprising the biological target and a paramagnetically labeled derivative of a first ligand to define a paramagnetic zone;
   b) preparing second NMR spectra of a second complex comprising the biological target and a second ligand;
   c) analyzing the spectra to determine whether the second ligand binds to the biological target within the paramagnetic zone of the paramagnetically labeled derivative;
   d) deducing a relative three-dimensional orientation of the first and second ligands when bound to the biological target;
   e) deducing a distance of separation of the first and second ligands when bound to the biological target; and
   f) selecting or preparing a compound that contains the first and second ligands substantially in the relative orientation and distance, wherein the binding affinity of the first ligand for the biological target is improved.
2) The method of claim 1 wherein step (c) is performed by:
   a) identifying peaks on the first NMR spectra that are perturbed by the paramagnetic label; and
   b) determining whether the second ligand perturbs peaks on the second NMR spectra that are also perturbed by the paramagnetic label.
3) The method of claim 1 wherein the first complex further comprises the second ligand, and step (c) is performed by determining whether the paramagnetically labeled derivative of the first ligand perturbs peaks associated with the second ligand.
4) The method of claim 3 further comprising, before step (c), preparing third NMR spectra of a mixture of the paramagnetically labeled derivative of the first ligand and the second ligand in the absence of the biological target.
5) (CANCELLED)

6) The method of claim 1 wherein the distance of separation is determined as a function of the loss of intensity for NMR resonances from the second ligand.

7) The method of claim 1 wherein the three dimensional orientation is deduced by producing a field ordered state in a medium comprising the biological target, the first ligand, and the second ligand, and analyzing dipolar couplings within the first and second ligands.

8) The method of claim 7 wherein the field ordered state is produced by an aqueous dispersion of lipid bicelles having complementary charges to the biological target.

9) The method of claim 7 wherein the field ordered state is produced by an aqueous dispersion of bacteriophage having a domain of the biological target in the outer coat.

10) The method of claim 1 wherein the paramagnetic label is a nitroxide or metal chelate.

11) The method of claim 1 wherein the first and second NMR spectra are two dimensional heteronuclear single quantum coherence spectra.

12) The method of claim 1 wherein the biological target is isotopically labeled.

13) The method of claim 1 wherein the biological target is a protein, and NMR resonances from the protein are not assigned to a sequence of the protein.

14) The method of claim 1 wherein the biological target is a protein, and the three dimensional conformation of the protein is unknown.

15) The method of claim 1 wherein the first ligand is an oligosaccharide, and the biological target is a protein.

16) The method of claim 1 wherein the second complex comprises the first ligand, or a paramagnetically labeled derivative thereof.

17) (CANCELLED)

18) (CANCELLED)

19) A method for improving the binding affinity of ligands for biological targets comprising:

a) preparing first NMR spectra of a first complex comprising a biological target, a paramagnetically labeled derivative of a first ligand, and a second ligand;

b) preparing second NMR spectra of a second complex comprising the biological target and either the second ligand or the paramagnetically labeled derivative of the first ligand;

c) preparing third NMR spectra of a mixture of the paramagnetically labeled derivative of the first ligand and the second ligand in the absence of the biological target;

d) analyzing the spectra to determine whether the paramagnetically labeled derivative of the first ligand perturbs peaks associated with the second ligand;

e) deducing a relative three-dimensional orientation of the first and second ligands when bound to the biological target;

f) deducing a distance of separation of the first and second ligands when bound to the biological target; and

g) selecting or preparing a compound that contains the first and second ligands substantially in the relative orientation and distance, wherein the binding affinity of the first ligand for the biological target is improved.

20) The method of claim 19 further comprising:

a) deducing from the NMR spectra the distance between the first and second ligands when bound to the biological target;

b) deducing from the NMR spectra the relative three dimensional orientation of the first and second ligand when bound to the biological target; and

c) selecting or preparing a hybrid ligand that contains the first and second ligands covalently linked substantially at the bond distance and relative orientation deduced in steps (a) and (b).

APPENDIX 2

EVIDENCE APPENDIX

Attached hereto are copies of the evidence entered by the Examiner and relied upon by appellant in the appeal. This evidence was cited by the Examiner at the following points during prosecution of this application:

Johnson et al., J. Mol. Biol. (1999) 287:609-625, cited by the Examiner in Office Actions mailed February 25, 2004 and October 26, 2004.

Bolon et al., J. Mol. Biol. (1999) 293:107-115, cited by the Examiner in an Office Action mailed October 26, 2004.

APPENDIX 3

RELATED PROCEEDINGS APPENDIX

None

APPENDIX 1

CLAIMS APPENDIX

1) A method for improving the binding affinity of a ligand for a biological target comprising:
   a) preparing first NMR spectra of a first complex comprising the biological target and a paramagnetically labeled derivative of a first ligand to define a paramagnetic zone;
   b) preparing second NMR spectra of a second complex comprising the biological target and a second ligand;
   c) analyzing the spectra to determine whether the second ligand binds to the biological target within the paramagnetic zone of the paramagnetically labeled derivative;
   d) deducing a relative three-dimensional orientation of the first and second ligands when bound to the biological target;
   e) deducing a distance of separation of the first and second ligands when bound to the biological target; and
   f) selecting or preparing a compound that contains the first and second ligands substantially in the relative orientation and distance, wherein the binding affinity of the first ligand for the biological target is improved.
2) The method of claim 1 wherein step (c) is performed by:
   a) identifying peaks on the first NMR spectra that are perturbed by the paramagnetic label; and
   b) determining whether the second ligand perturbs peaks on the second NMR spectra that are also perturbed by the paramagnetic label.
3) The method of claim 1 wherein the first complex further comprises the second ligand, and step (c) is performed by determining whether the paramagnetically labeled derivative of the first ligand perturbs peaks associated with the second ligand.
4) The method of claim 3 further comprising, before step (c), preparing third NMR spectra of a mixture of the paramagnetically labeled derivative of the first ligand and the second ligand in the absence of the biological target.
5) (CANCELLED)

6) The method of claim 1 wherein the distance of separation is determined as a function of the loss of intensity for NMR resonances from the second ligand.

7) The method of claim 1 wherein the three dimensional orientation is deduced by producing a field ordered state in a medium comprising the biological target, the first ligand, and the second ligand, and analyzing dipolar couplings within the first and second ligands.

8) The method of claim 7 wherein the field ordered state is produced by an aqueous dispersion of lipid bicelles having complementary charges to the biological target.

9) The method of claim 7 wherein the field ordered state is produced by an aqueous dispersion of bacteriophage having a domain of the biological target in the outer coat.

10) The method of claim 1 wherein the paramagnetic label is a nitroxide or metal chelate.

11) The method of claim 1 wherein the first and second NMR spectra are two dimensional heteronuclear single quantum coherence spectra.

12) The method of claim 1 wherein the biological target is isotopically labeled.

13) The method of claim 1 wherein the biological target is a protein, and NMR resonances from the protein are not assigned to a sequence of the protein.

14) The method of claim 1 wherein the biological target is a protein, and the three dimensional conformation of the protein is unknown.

15) The method of claim 1 wherein the first ligand is an oligosaccharide, and the biological target is a protein.

16) The method of claim 1 wherein the second complex comprises the first ligand, or a paramagnetically labeled derivative thereof.

17) (CANCELLED)

18) (CANCELLED)

19) A method for improving the binding affinity of ligands for biological targets comprising:

a) preparing first NMR spectra of a first complex comprising a biological target, a paramagnetically labeled derivative of a first ligand, and a second ligand;

b) preparing second NMR spectra of a second complex comprising the biological target and either the second ligand or the paramagnetically labeled derivative of the first ligand;

c) preparing third NMR spectra of a mixture of the paramagnetically labeled derivative of the first ligand and the second ligand in the absence of the biological target;

d) analyzing the spectra to determine whether the paramagnetically labeled derivative of the first ligand perturbs peaks associated with the second ligand;

e) deducing a relative three-dimensional orientation of the first and second ligands when bound to the biological target;

f) deducing a distance of separation of the first and second ligands when bound to the biological target; and

g) selecting or preparing a compound that contains the first and second ligands substantially in the relative orientation and distance, wherein the binding affinity of the first ligand for the biological target is improved.

20) The method of claim 19 further comprising:

a) deducing from the NMR spectra the distance between the first and second ligands when bound to the biological target;

b) deducing from the NMR spectra the relative three dimensional orientation of the first and second ligand when bound to the biological target; and

c) selecting or preparing a hybrid ligand that contains the first and second ligands covalently linked substantially at the bond distance and relative orientation deduced in steps (a) and (b).

APPENDIX 2

EVIDENCE APPENDIX

Attached hereto are copies of the evidence entered by the Examiner and relied upon by appellant in the appeal. This evidence was cited by the Examiner at the following points during prosecution of this application:

Johnson et al., J. Mol. Biol. (1999) 287:609-625, cited by the Examiner in Office Actions mailed February 25, 2004 and October 26, 2004.

Bolon et al., J. Mol. Biol. (1999) 293:107-115, cited by the Examiner in an Office Action mailed October 26, 2004.

doi:10.1006/jmbi.1999.2827 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 287, 609–625

JMB

The Cellulose-binding Domains from *Cellulomonas fimi* β-1,4-Glucanase CenC Bind Nitroxide Spin-labeled Cellooligosaccharides in Multiple Orientations

Philip E. Johnson, Emmanuel Brun, Lloyd F. MacKenzie, Stephen G. Withers and Lawrence P. McIntosh*

Protein Engineering Network of Centres of Excellence and Department of Chemistry and Department of Biochemistry and Molecular Biology University of British Columbia Vancouver, British Columbia V6T 1Z3, Canada

**Corresponding author*

The N-terminal cellulose-binding domains CBD_{N1} and CBD_{N2} from *Cellulomonas fimi* cellulase CenC each adopt a jelly-roll β-sandwich structure with a cleft into which amorphous cellulose and soluble cellooligosaccharides bind. To determine the orientation of the sugar chain within these binding clefts, the association of TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl-4-yl) spin-labeled derivatives of cellotriose and cellotetraose with isolated CBD_{N1} and CBD_{N2} was studied using heteronuclear ^{1}H-^{15}N NMR spectroscopy. Quantitative binding measurements indicate that the TEMPO moiety does not significantly perturb the affinity of the cellooligo-saccharide derivatives for the CBDs. The paramagnetic enhancements of the amide $^{1}H^{N}$ longitudinal (ΔR_1) and transverse (ΔR_2) relaxation rates were measured by comparing the effects of TEMPO-cellotetraose in its nitroxide (oxidized) and hydroxylamine (reduced) forms on the two CBDs. The bound spin-label affects most significantly the relaxation rates of amides located at both ends of the sugar-binding cleft of each CBD. Similar results are observed with TEMPO-cellotriose bound to CBD_{N1}. This demonstrates that the TEMPO-labeled cellooligosaccharides, and by inference strands of amorphous cellulose, can associate with CBD_{N1} and CBD_{N2} in either orientation across their β-sheet binding clefts. The ratio of the association constants for binding in each of these two orientations is estimated to be within a factor of five to tenfold. This finding is consistent with the approximate symmetry of the hydrogen-bonding groups on both the cellooligosaccharides and the residues forming the binding clefts of the CenC CBDs.

Keywords: protein-carbohydrate interaction; NMR; spin label; glycosynthase

Introduction

The β-1,4-glucanase CenC from *Cellulomonas fimi* is modular in structure and function (Coutinho *et al.*, 1992, 1991). Its central catalytic domain is connected by a short linker to two tandem N-terminal cellulose-binding domains, CBD_{N1} and CBD_{N2}(Tomme *et al.*, 1995). In contrast to most CBDs which bind crystalline cellulose, the homologous family IV CBD_{N1} and CBD_{N2} from CenC are unique in their specificity for amorphous cellulose and soluble cellooligosaccharides. As part of a col-

Present address: P. E. Johnson, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, MD 21250, USA.

Abbreviations used: CBD, cellulose-binding domain; CBD_{Cex}, the cellulose-binding domain from the mixed cellulase/xylanase *Cellulomonas fimi* Cex; CBD_{N1}, the N-terminal cellulose-binding domain from *Cellulomonas fimi* β-1,4-glucanase CenC; CBD_{N2}, the cellulose-binding domain from *Cellulomonas fimi* β-1,4-glucanase CenC following CBD_{N1} in sequence; CBD_{N1N2}, the tandem cellulose-binding domains from *Cellulomonas fimi* β-1,4-glucanase CenC; HSQC, heteronuclear single quantum correlation; NOE, nuclear Overhauser effect; pH*, the observed pH meter reading without correction for isotope effects; TEMPO, 2,2,6,6-tetramethylpiperidine-1-oxyl-4-yl; TEMPO-Glc_3, TEMPO-labeled cellotriose; TEMPO-Glc_4, TEMPO-labeled cellotetraose.

E-mail address of the corresponding author: mcintosh@otter.biochem.ubc.ca

0022-2836/99/130609–17 $30.00/0

laborative effort to exploit these domains for use in biotechnology, we have used NMR spectroscopy and calorimetry to provide a structural and thermodynamic foundation for understanding the distinct binding properties of CBD_{N1} and CBD_{N2} (Creagh *et al.*, 1998; Johnson *et al.*, 1996a,b, 1998; Tomme *et al.*, 1996a).

CBD_{N1} (Johnson *et al.*, 1996b) and CBD_{N2} are composed of two five-stranded antiparallel (E.B., P.E.J. & L.P.M., unpublished results) β-sheets that fold into very similar jelly-roll sandwich structures. As demonstrated by NMR chemical shift perturbations and the observation of intermolecular protein-sugar NOEs, a single cellooligosaccharide molecule binds to these CBDs within a groove or cleft that runs across one β-sheet face of each protein domain (Johnson *et al.*, 1996a,b). The presence of a cleft readily explains the binding specificity of these CBDs. Soluble cellooligosaccharides and single polysaccharide chains, as might be encountered in the amorphous regions of cellulose, bind within this cleft, whereas flat crystalline arrays of cellulose are excluded. In agreement with the observation that the affinity of CBD_{N1} for cellooligosaccharides increases in the order cellotriose < cellotetraose < cellopentaose ~ cellohexaose, the sugar binding sites in the CenC CBDs can each be spanned by approximately five glycosyl units (Johnson *et al.*, 1996a,b; Tomme *et al.*, 1996a).

In parallel with these structural studies, a detailed calorimetric analysis of the association of CBD_{N1} with polysaccharides was also conducted (Creach *et al.*, 1998; Tomme *et al.*, 1996a). CBD_{N1} was found to bind regenerated (phosphoric acid swollen) cellulose and other soluble β-1,4-linked polymers of glucose, such as hydroxyethyl cellulose and barley and oat-β-glucan, with affinities equal to those measured for cellopentaose and cellohexaose. Formation of the CBD_{N1}-sugar complexes is favored enthalpically, indicating that hydrogen bonding and van der Waals interactions provide the primary driving force for the binding event. Inspection of the structures of CBD_{N1} and CBD_{N2} reveals that their binding clefts consist of a central strip of hydrophobic side-chains, flanked on both sides by polar residues. This led us to propose that the pyranose rings of cellooligosaccharides, and by inference single polysaccharide chains in regions of amorphous cellulose, are stacked against the hydrophobic strip, while the flanking hydrophilic residues provide hydrogen bonds to the equatorial hydroxyl groups of the sugar (Johnson *et al.*, 1996a,b). Such interactions are often observed with carbohydrate-binding proteins (Quiocho, 1989, 1993; Vyas, 1991).

The structures of CBD_{N1} and CBD_{N2} were calculated using NMR data collected for these protein domains in the presence of saturating concentrations of cellotetraose or cellopentaose, respectively (Johnson *et al.*, 1996b). Numerous intermolecular nuclear Overhauser effects (NOEs) between the $^{13}C/^{15}N$-labeled proteins and the unlabeled cellooligosaccharides were observed in isotope edited/filtered NOESY experiments (Otting & Wüthrich, 1990). However, the NMR spectra of the cellooligosaccharides are highly degenerate, thus preventing the unambiguous assignment of these NOEs to specific interactions between sugar and protein protons. This precluded the direct determination of the structures of the CBD-cellooligosaccharide complexes.

As a first step towards achieving this goal, we have focussed on defining the orientation of a cellulose chain within the binding clefts of CBD_{N1} and CBD_{N2}. Exploiting a novel glycosynthase technology (Mackenzie *et al.*, 1998), we have prepared derivatives of cellotriose and cellotetraose with a 2,2,6,6-tetramethylpiperidine-1-oxy-4-yl (TEMPO) spin-label covalently attached to the reducing end of the sugar (Figure 1). The use of such paramagnetic relaxation probes has a long history in the study of biological macromolecules by NMR spectroscopy (Kosen, 1989). The nitroxide moiety contains an unpaired electron that provides an efficient mechanism for the relaxation of neighboring nuclei *via* dipolar coupling. Due to the magnitude of the electron magnetic moment, this interaction extends to over 20 Å. In contrast, proton-proton NOEs are limited to separations of less than ~5 Å. Thus, these TEMPO-labeled sugars can be utilized to obtain long-range distance information about the CBD-cellooligosaccharide complexes. In practice, the excellent dispersion of the signals from the amide ^{1}H-^{15}N groups, combined with the high degree of sensitivity of the HSQC experiment (Cavanagh *et al.*, 1996), allows the effect of the spin-label cellooligosaccharides on

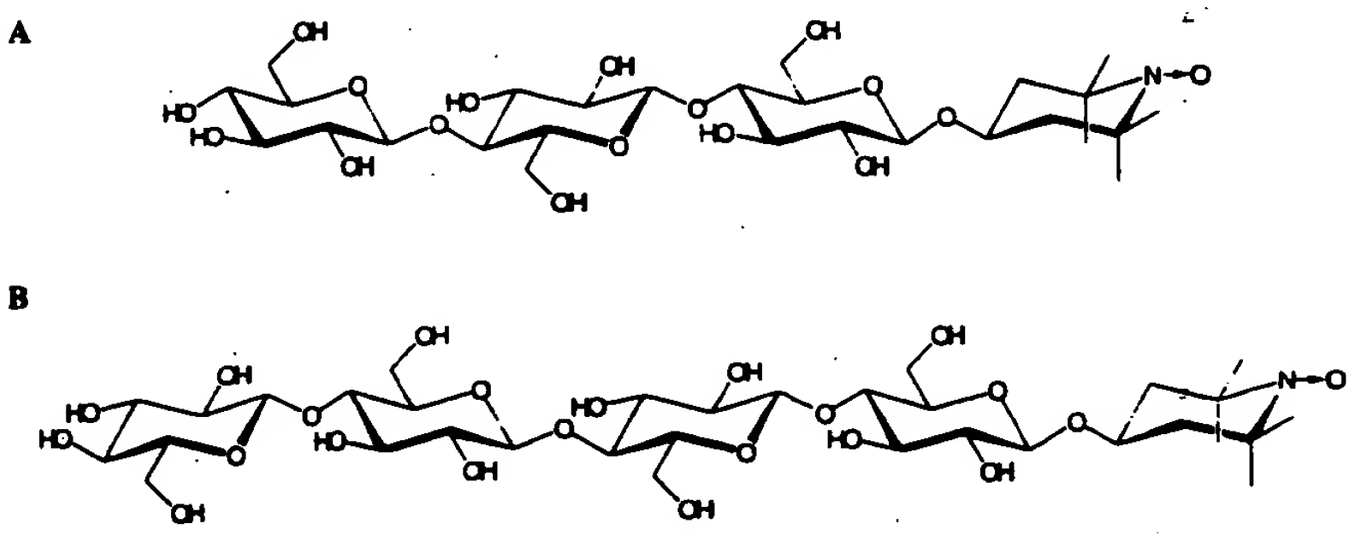


Figure 1. Chemical structures of the spin-labeled glycosides: (a) TEMPO-Glc_3 and (b) TEMPO-Glc_4. Reduction of the paramagnetic nitroxide (N-O) moiety produces the diamagnetic hydroxylamine (N-OH) derivatives of these glycosides.

each non-proline residue in ^{15}N-labeled CenC CBDs to be evaluated. Similar studies using spin-labeled ligands and heteronuclear NMR methods have been reported by several groups (Gellespie & Shortle, 1997a,b; Zhao *et al.*, 1997; Kleerekoper *et al.*, 1995; Yu *et al.*, 1994). Using this approach, we demonstrate that the modified sugars bind to CBD_{N1} and CBD_{N2} in both possible orientations across their β-sheet binding clefts. This result is consistent with the approximate symmetry of the cellooligosaccharides and the positions of the side-chains forming the binding clefts of the two protein domains, as well as ^{15}N relaxation measurements which indicate that residues within the binding cleft of CBD_{N1} remain conformationally flexible in the presence of these sugars (P.E.J., E.B. & L.P.M., unpublished results).

Results

Preparation of TEMPO-labeled cellooligosaccharides

Syntheses of the nitroxide spin-labeled cellotrioside (TEMPO-Glc_3) and cellotetraoside (TEMPO-Glc_4) were accomplished in a two-step chemoenzymatic process, thereby avoiding many of the difficulties inherent in the preparation of oligosaccharide derivatives by chemical methods alone. A TEMPO-labeled glucoside was first synthesized *via* standard Koenigs-Knorr methodology, essentially according to a published procedure (Gnewuch & Sosnovsky, 1986; Plessas & Goldstein, 1981). This product was characterized by mass and, to a limited extent, ^{1}H-NMR spectroscopy. Full NMR spectroscopic characterization was performed on the catalytically hydrogenated (reduced) and acetylated hydroxylamine derivative that no longer bears a free radical. Conversion of the TEMPO-glucoside to TEMPO-Glc_3 and Glc_4 was accomplished by successive transfer of glucose residues from α-glucosyl fluoride using a mutant *Agrobacterium* sp. β-glucosidase (glycosynthase) in which the catalytic nucleophile, Glu358, was replaced by an alanine residue (Mackenzie *et al.*, 1998). This approach exploits the well-known transglycosylation reaction catalyzed by retaining glycosidases, yet has the advantage that the mutant enzyme cannot hydrolyze the end products, thereby resulting in good yields of ~70% per coupling. The products were purified by HPLC and their identities confirmed by ionspray mass spectrometry, and after reduction and acetylation, by ^{1}H-NMR spectroscopy.

Binding of TEMPO-labeled cellooligosaccharides to CBD_{N1} and CBD_{N2}

We have demonstrated previously, using NMR spectroscopy and isothermal titration calorimetry, that the monomeric CBD_{N1} binds soluble cellooligosaccharides with a stoichiometry of 1:1 (Johnson *et al.*, 1996a; Tomme *et al.*, 1996a). Similar results have been obtained with CBD_{N2} (data not shown). Continuing this approach, the interactions of the TEMPO-labeled cellooligosaccharides with CBD_{N1} and CBD_{N2} were analyzed quantitatively by monitoring the $^1H^N$ and ^{15}N chemical shifts of the proteins upon titration with these soluble sugars. The spectral changes resulting from the addition of TEMPO-Glc_3 and TEMPO-Glc_4 are similar to those observed with the corresponding unlabeled cellooligosaccharides, indicating that these ligands all bind in the regime of fast exchange on the chemical shift timescale to the same regions in CBD_{N1} and CBD_{N2} and with the same 1:1 stoichiometry (data not shown). Apparent or macroscopic equilibrium association constants were obtained by fitting the titration data to the Langmuir isotherm describing the binding of one ligand molecule to a single protein site (Figure 2).

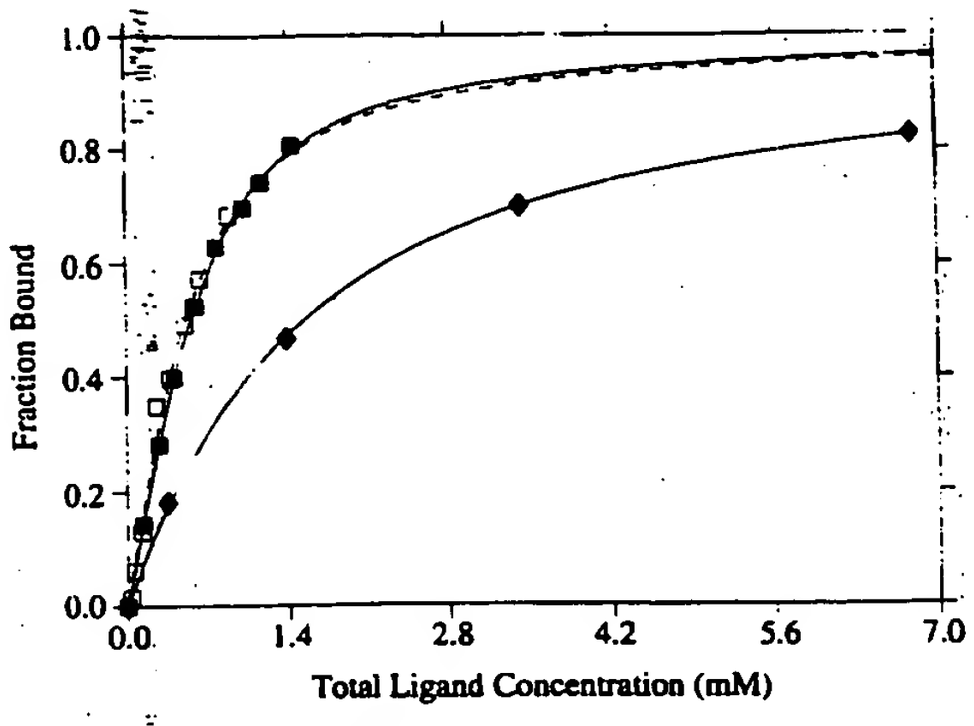


Figure 2. The association constants of the CenC CBDs for TEMPO-labeled cellooligosaccharides were measured using ^{1}H-^{15}N NMR spectroscopy. Shown are the normalized chemical shifts changes, or fraction bound, of the amide proton of Asn81 in CBD_{N1} upon titration with TEMPO-Glc_3 (filled diamonds) and TEMPO-Glc_4 (filled squares), and the nitrogen atom of Val197 in CBD_{N2} upon addition of TEMPO-Glc_4 (open squares). The solid and broken lines represent the best fits of the data to the Langmuir isotherm describing the binding of one ligand to a single protein site on CBD_{N1} or CBD_{N2}, respectively. The total chemical shift changes for Asn81 and Val197 were extrapolated to be 0.1 ppm ($^1H^N$) and 0.76 ppm (^{15}N), respectively.

TEMPO-Glc_3 binds to CBD_{N1} with an apparent K_a of 690(±260) M^{-1}. This is marginally higher than the value of 180(±60) M^{-1} measured previously for unlabeled cellotriose (Johnson *et al.*, 1996a). The apparent K_a values of TEMPO-Glc_4 for CBD_{N1} and CBD_{N2} are 4200(±400) M^{-1} and 3600(±2200) M^{-1}, respectively, while the association constants of CBD_{N1} and CBD_{N2} for unlabeled cellotetraose are 4200(±720) M^{-1} (Johnson *et al.*, 1996a) and 7000(±650) M^{-1} (data not shown). The relatively large errors in the K_a values measured for the TEMPO-labeled cellooligosaccharides reflect the fact that the nitroxide efficiently broadens the

signals from many amide groups within the binding cleft to the point that they are no longer observable. In general, it is these amide groups that show the largest chemical shift changes upon addition of sugar to the protein and would otherwise be best suited for fitting to the binding isotherm equation. Within the error of the data, we conclude that the TEMPO moiety does not significantly perturb the affinity of the CenC CBDs for these cellooligosaccharides. By way of comparison, increases in affinity of ~10 to 20-fold result from the lengthening of cellotriose or cellotetraose by one additional glycosyl unit to form cellotetraose or cellopentaose, respectively (Tomme *et al.*, 1996a).

TEMPO-labeled cellooligosaccharides bind CBD_{N1} and CBD_{N2} in multiple orientations

In addition to perturbing the chemical shifts of amide groups in CBD_{N1} and CBD_{N2}, the binding of TEMPO-Glc_3 and Glc_4 causes a decrease in the intensities of the signals from many of these groups due to paramagnetic relaxation enhancement. This is readily seen by comparing the ^1H-^{15}N HSQC spectra of two proteins complexed with the oxidized and reduced cellotetraose derivative

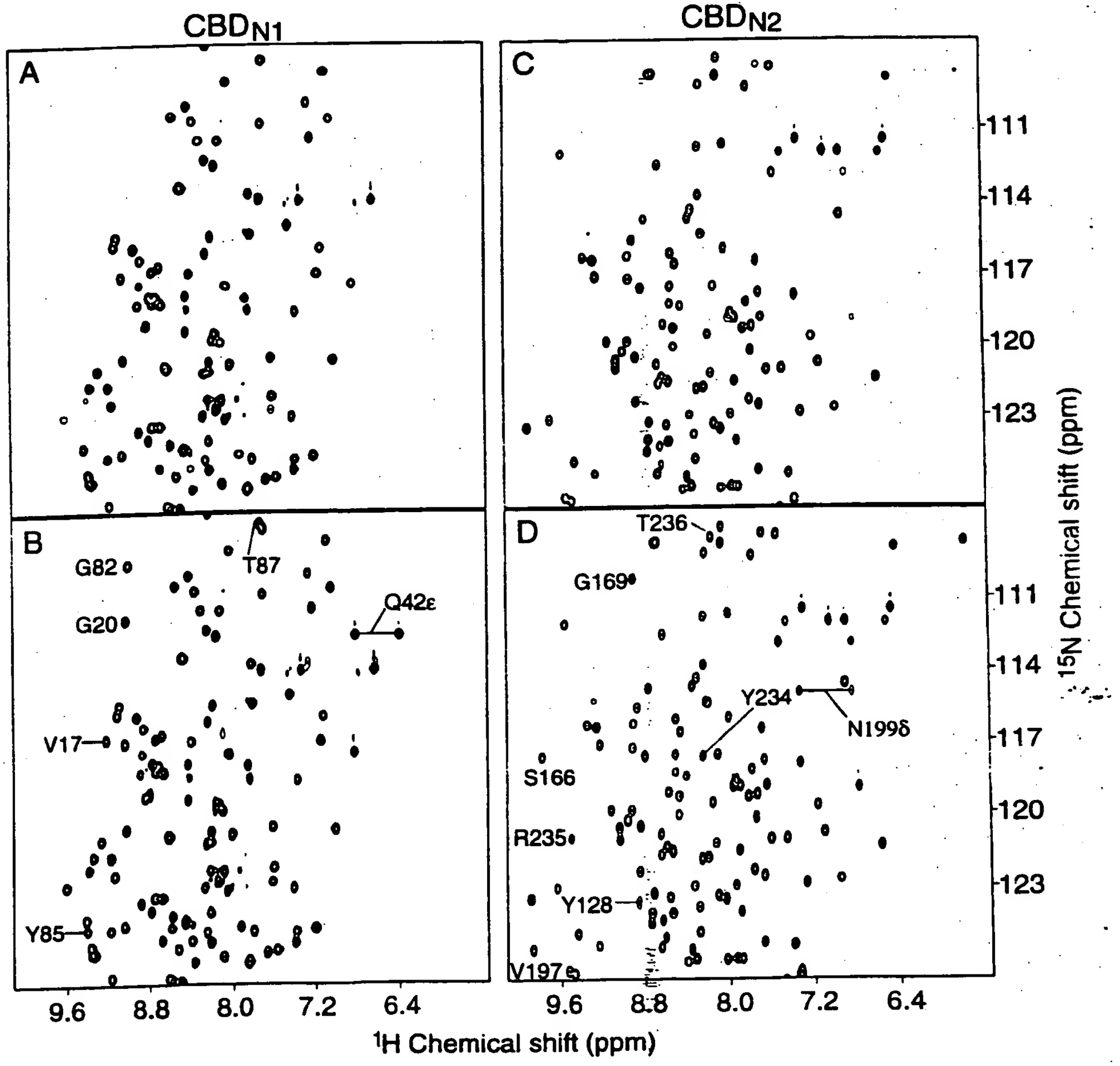


Figure 3. Portions of the non-sensitivity enhanced ^1H-^{15}N HSQC spectra of (a), (b) CBD_{N1} and (c), (d) CBD_{N2} bound to (a), (c) oxidized or (b), (d) reduced TEMPO-Glc_4. The final ligand-to-protein ratios of 2.2:1 and 1.4:1 correspond to ~80% and 65% saturation of CBD_{N1} and CBD_{N2}, respectively. Peaks from amides that "disappear" because of paramagnetic relaxation due to the nearby unpaired electron in the nitroxide-labeled cellotetraose are identified.

(Figure 3). Qualitatively, it is clear that a number of resonances present in the spectra of proteins bound to the reduced hydroxylamine form of the sugar are "absent" in the spectra recorded with the oxidized nitroxide form, i.e. their intensities fall below a very low contour level. Thus, these amide groups must be close ($\leqslant$11 Å; *vide infra*) to the free electron in the nitroxide moiety.

The amide protons whose resonances are absent in the ^1H-^{15}N HSQC spectra of the complexes of TEMPO-Glc_4 with CBD_{N1} and CBD_{N2} can be classified into two groups that are located on the opposite ends of the binding clefts of these proteins. (As shown in Figure 5, the binding cleft of each CBD lies across an antiparallel β-sheet A formed by five strands, arranged as A1-A2-A5-A3-A4.) In the case of CBD_{N1}, the first group includes Val17, Ala18, Tyr19 and Gly20 from β-strand A1, while the second group comprises Tyr85, Gly86 and Thr87 from β-strand A4 and the loop between A3 and A4. Similarly, for CBD_{N2}, the first group is formed by Ser166, Leu167, Tyr168, Gly169 from β-strand A1, while the second group includes of Tyr234, Arg235 and Thr236 from β-strand A4 and the preceding loop. In contrast, peaks from amide protons in the central β-strand A5 of each CBD are present in the ^1H-^{15}N HSQC spectra. Inspection of the tertiary structures of CBD_{N1} and CBD_{N2} reveals that the outermost strands, A1 and A4, are separated by ~20 Å. The observation that the amide protons, which are most strongly perturbed by the nitroxide spin label, are located in both of these outer β-strands cannot be accounted for by the binding of TEMPO-Glc_4 to the CBDs in a single orientation. The simplest explanation of this result is that TEMPO-Glc_4 associates with CBD_{N1} and CBD_{N2} in at least two distinct orientations, such that the nitroxide moiety lies at either edge of their sugar binding clefts.

The peaks from amide protons in β-strands A1 and A4 that are absent in the spectrum of CBD_{N1} bound to TEMPO-Glc_4 also disappear upon binding TEMPO-Glc_3 (see Supplementary Material). In addition, signals from several amide protons, including those of Thr21 in strand A1 and Ala41, $Gln42^{\varepsilon}$, Tyr43, Val48, Asn50, and Gly51 in strand A2, are weak in the spectrum of CBD_{N1} complexed with TEMPO-Glc_4 and absent in the presence of TEMPO-Glc_3. This qualitative distinction may reflect a slight difference in the positions of the two sugars when bound to CBD_{N1}. Several protons, including the indole $H^{\varepsilon 1}$ of Trp16 and the amide H^N of Leu146, which are not in the binding cleft of CBD_{N1}, are also affected by TEMPO-Glc_3. This is attributed to non-specific effects arising from the addition of a 22-fold molar excess of the nitroxide-labeled sugar to the protein sample.

Measurement of ΔR_1 and ΔR_2 relaxation enhancements

A qualitative inspection of the ^1H-^{15}N HSQC spectra of the two CBD-TEMPO-Glc_4 complexes provides immediate insights into the binding orientations of the sugars. Further details regarding the positions of the bound nitroxide moiety can be gained by a quantitative analysis of the effects of the spin label on the relaxation properties of the two CBDs. To reduce the possibility of non-specific interactions, CBD_{N1} and CBD_{N2} were studied at saturation levels of approximately 80% and 65%, respectively (Figure 2).

The enhancements of the longitudinal (ΔR_1) and transverse (ΔR_2) relaxation rates of CBD_{N1} and CBD_{N2} due to complexation with TEMPO-Glc_4 are presented in Figures 4 and 5. Data for CBD_{N1} bound to TEMPO-Glc_3 are provided as Supplementary Material. The ΔR_1 values were determined from the differences in the T_1 lifetimes, measured using a ^1H-^{15}N HSQC sequence as a read-out of a non-selective inversion-recovery sequence (Cavanagh *et al.*, 1996), for the proteins in the presence of the nitroxide and hydroxylamine derivatives of cellotetraose. The ΔR_2 values were obtained using two complementary methods. In the first, the effect of the spin label on the proton transverse relaxation ($\Delta R_{2,Vol}$) of the $^1H^N$ during the INEPT and reverse-INEPT period of the HSQC pulse sequence was measured by comparing the total volume of its cross-peak in spectra recorded with the oxidized and reduced sugar. In the second, $\Delta R_{2,LW}$ was determined from the change in the proton line-width of the amide measured with the two forms of the modified cellotetraose. The methods for determining ΔR_1 and ΔR_2 assume that contributions to the peak volumes and linewidths due to scalar couplings and relaxation processes other than that mediated by the free radical remain constant, and that the oxidized and reduced glycosides interact identically with the CBDs. The lack of any significant amide $^1H^N$ or ^{15}N chemical shift perturbation upon reduction of the TEMPO group provides strong support for these assumptions (Figure 3).

Following these approaches, 101 ΔR_1, 64 $\Delta R_{2,Vol}$, and 99 $\Delta R_{2,LW}$ values were obtained for CBD_{N1} and 113 ΔR_1, 44 $\Delta R_{2,Vol}$, and 118 $\Delta R_{2,LW}$ values for CBD_{N2} (Figure 4). Rate enhancements were not measurable for the remaining non-proline residues for two reasons, besides spectral overlap. First, in the case of those amide protons that are closest to the nitroxide moiety, severe line-broadening rendered their signals undetectable in the spectra recorded with oxidized TEMPO-Glc_4. For these residues, we estimate $\Delta R_1 > 2.5\ s^{-1}$ and $\Delta R_2 > 80\ s^{-1}$. Second, in the case of those amide protons furthest from the nitroxide group, changes in their relaxation rates were too small to measure reliably. In particular, dilution effects due to the addition of ascorbic acid and subsequent pH adjustments caused the volumes of the peaks in the ^1H-^{15}N HSQC spectra of the proteins recorded with the reduced sugar to decrease slightly. This led to small apparent negative $\Delta R_{2,Vol}$ values for many

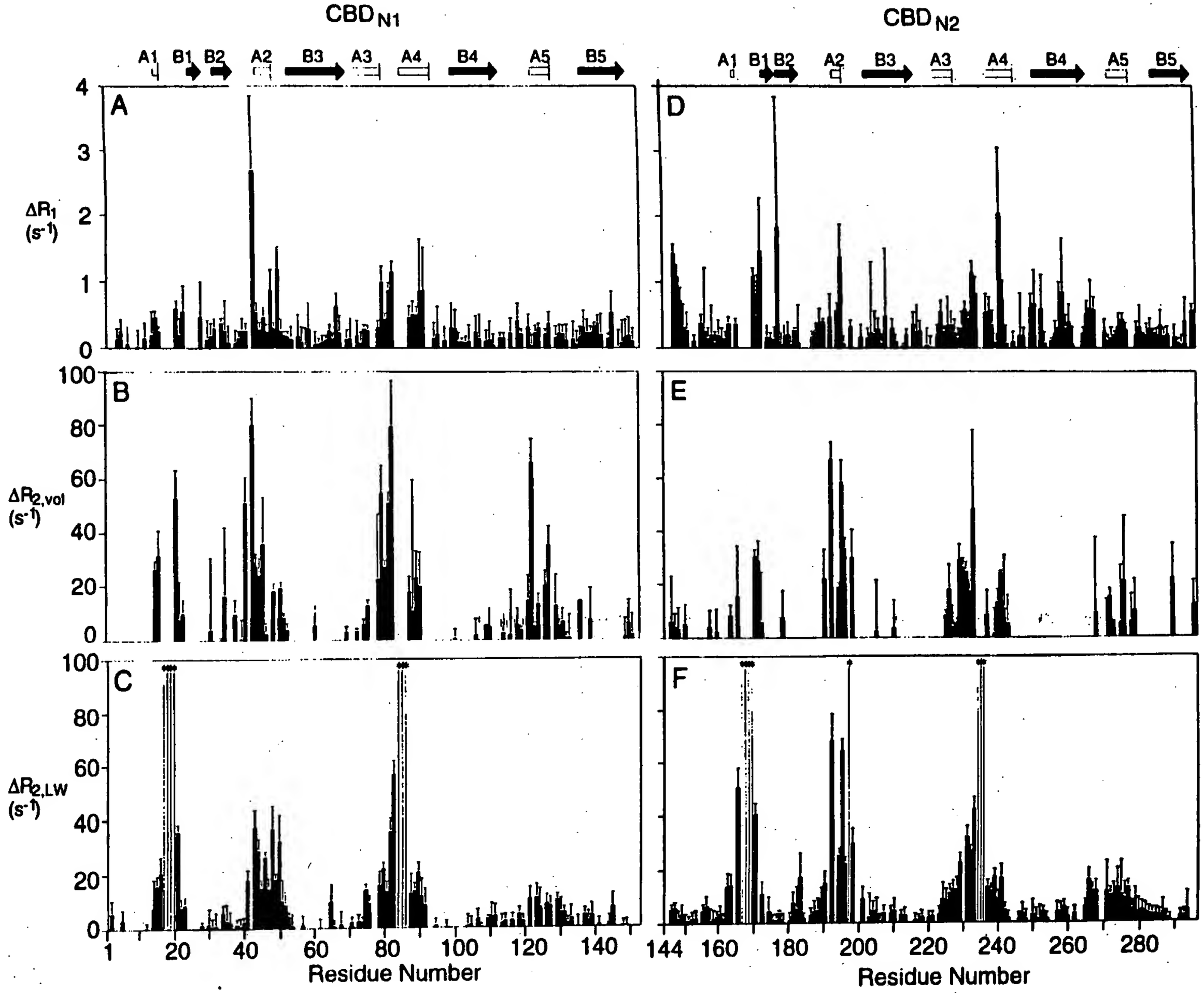


Figure 4. The enhancement of the amide $^1H^N$ longitudinal (ΔR_1) and transverse ($\Delta R_{2,Vol}$ and $\Delta R_{2,LW}$) rates resulting from the binding of TEMPO-Glc_4 by (a), (b), (c) CBD_{N1} and (d), (e), (f) CBD_{N2}. Only positive rate enhancements are shown. Residues whose signals disappear completely due to the nitroxide spin label are identified in (c) and (f) with an asterisk. The locations of the β-strands forming sheets A and B in the two proteins are indicated by the open and filled arrows, respectively.

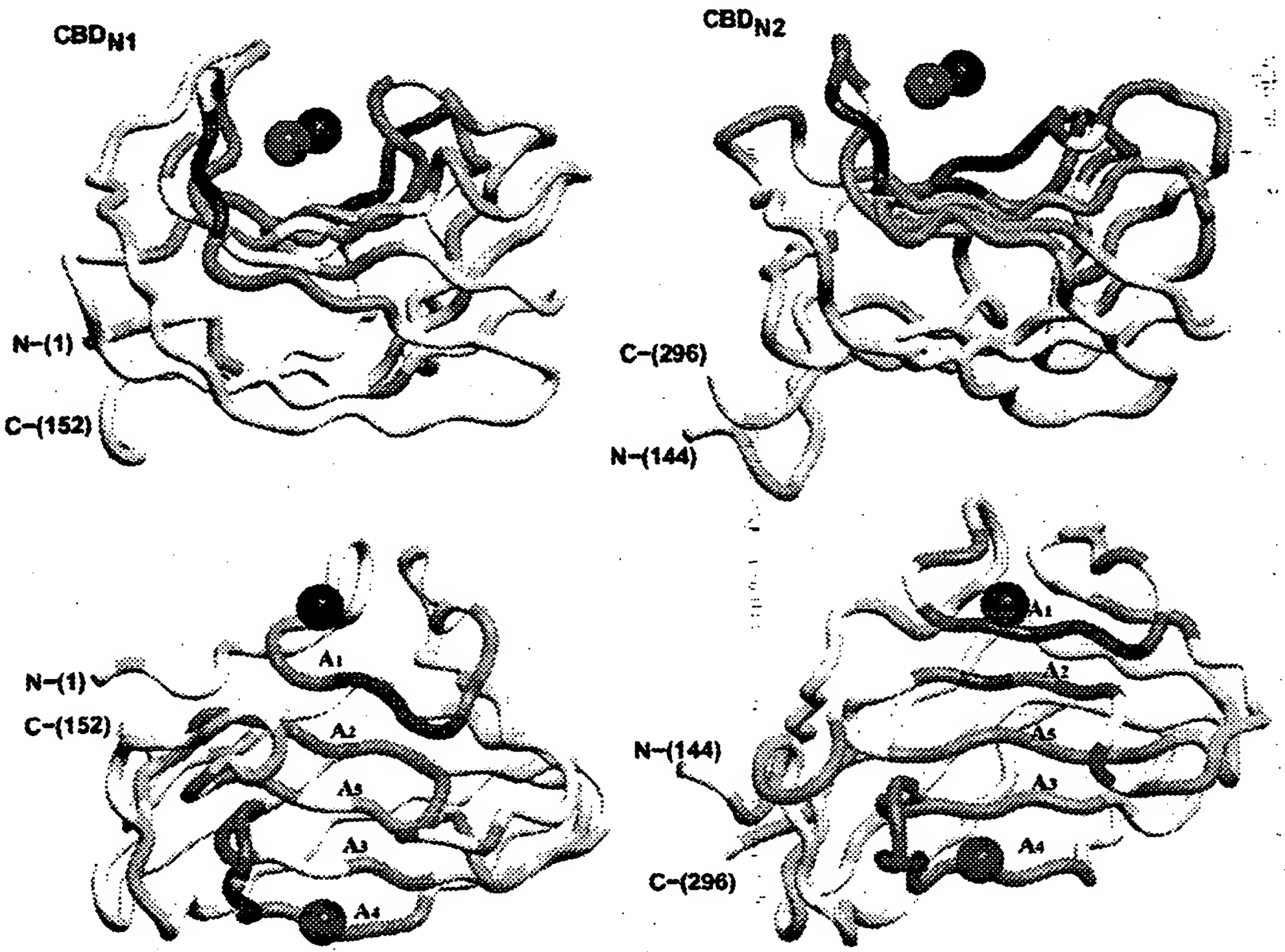


Figure 5. Paramagnetic relaxation studies demonstrate that TEMPO-Glc_4 binds to each CenC CBD in at least two orientations such that the nitroxide moiety can lie at either end of the binding cleft. Shown are ribbon diagrams of CBD_{N1} (left) and CBD_{N2} (right) positioned for the reader to look across (top) and down onto (bottom) the cellooligosaccharide binding groove that is formed by β-sheet A in each protein. The individual strands of this β-sheet are labeled. The ribbons are ramp-colored based upon $\Delta R_{2,LW}$ values ranging from 1 Hz (light yellow) to 80 Hz (dark orange). Residues whose $^1H^N$ signals disappear completely in the presence of TEMPO-Glc_4 are indicated in red ($\Delta R_{2,LW} > 80$ Hz). The proline residues and other residues for which no results were obtained are shown in white. The green and blue spheres indicate representative positions of the nitroxide groups calculated using an ensemble-averaging protocol in X-PLOR with measured r_{eff} as distance restraints. Note that the TEMPO-Glc_4 can occupy only one of these two orientations at any given time.

amides whose relaxation rates were not perturbed significantly by the spin label.

As seen in Figure 4, ΔR_1 and ΔR_2 show similar trends, yet the latter provide a more sensitive indicator of the effects of the spin-labeled sugars on CBD_{N1} and CBD_{N2}. In the cases of amide protons with resolved ^1H-^{15}N HSQC cross-peaks that experience moderate paramagnetic relaxation enhancements, the values of $\Delta R_{2,Vol}$ and $\Delta R_{2,LW}$ agree well (ratio = 1.7(±1)). However, a more complete data set could be obtained by measuring changes in line-width rather than volume due to the above mentioned problems associated with dilution effects and difficulties encountered with partially overlapping peaks.

Based on these measurements, it is immediately apparent that amide groups in four regions of CBD_{N1} and four regions of CBD_{N2} experience significant paramagnetic relaxation enhancements due to the bound TEMPO-Glc_4. These include Gly15 to Thr21, Tyr43 to Asn51, Gln80 to Asp90 and Glu122 to Ser133 in CBD_{N1} and Gly160 to Glu172, Trp192 to Asn199, Glu229 to Glu239 and Gly266 to Ala279 in CBD_{N2}. These residues lie within the five β-strands, A1 to A5, and adjacent loops that form the cellulose-binding clefts of each protein. Of these four regions, amide proteins in strands A1 and A4 that comprise the outer edges of the binding face have markedly larger ΔR_1 and ΔR_2 values than do those in the central strand A5. This verifies that TEMPO-Glc_4 binds CBD_{N1} and CBD_{N2} in multiple orientations such that the nitroxide spin label is located at either end of the binding cleft. In contrast, the relaxation rates of amide groups located on the opposite side of each protein relative to this cleft (β-strands B1 to B5) do not exhibit marked enhancements. Similar patterns are observed with TEMPO-Glc_3 bound to CBD_{N1} (Supplementary Material).

Calculation of electron-proton distances

Values of τ_c, the correlation time for the fluctuation of the electronic-nuclear dipole-dipole interaction, were calculated for individual residues in both CBD_{N1} and CBD_{N2} bound to TEMPO-Glc_4 using the measured ΔR_1 and $\Delta R_{2,Vol}$ or $\Delta R_{2,LW}$ rate enhancements. In the case of CBD_{N1}, τ_c values for 81 amide protons ranged from 0.1 ns to 3.6 ns, with a mean of 1.7(±0.5) ns. For CBD_{N2}, these ranged from 0.1 ns to 6.5 ns for 101 amide protons, with a mean of 1.9(±0.8) ns (Figure 6). The correlation time τ_c is dependent upon both the relaxation of the electron and motions of the electron-proton vector ($1/\tau_c = 1/\tau_S + 1/\tau_R$). In this equation, τ_S is the longitudinal relaxation time of the free electron and τ_R is the effective rotational correlation time of the electron-proton vector. The latter is dependent on the motional characteristics of the protein-ligand complex. Since τ_R is in the range of 10^{-8} to 10^{-9} s for most proteins studied by solution NMR methods, and τ_S is typically longer than 10^{-7} s for nitroxide radicals, τ_c is essentially equal to τ_R (Kosen, 1989). The correlation times for the global tumbling of CBD_{N1} and CBD_{N2} saturated with cellopentaose are approximately 7.4 ns, as determined by ^{15}N relaxation methods (P.E.J., E.B. & L.P.M., unpublished results). The shorter correlation times extracted from the paramagnetic relaxation enhancements may reflect conformational mobility of the bound TEMPO moiety. The large degree of variability in the τ_c values measured for the amide protons in CBD_{N1} and CBD_{N2} is attributed to compounding errors from the ΔR_1 and ΔR_2 measurements, as well as different orientation-dependent effects of internal mobility on the distance between the free electron and the $^1H^N$ of individual amide protons.

Using the τ_c values measured for each residue, the effective electron-proton distances, r_{eff}, between the TEMPO-Glc_4 nitroxide and the individual amide protons in CBD_{N1} and CBD_{N2} were calculated. A similar analysis of the data for the TEMPO-Glc_3 was not performed due to the occurrence of non-specific relaxation enhancement. The most complete set of relaxation data was obtained from linewidth measurements, and thus the distances reported in Figure 6 were determined using the $\Delta R_{2,LW}$ values; however, similar results were derived from the $\Delta R_{2,Vol}$ rate enhancements. Based on the data in Figure 6, we find that the measured values of r_{eff} range from 11 to 23 Å. Consistent with the previous discussion of the data presented in Figures 2 and 3, the amide protons in CBD_{N1} and CBD_{N2} closest to the nitroxide group are located in β-strands A1 and A4 on the binding faces of the proteins, whereas those that are furthest form strands B1 to B5.

Structure calculations

The positions of the nitroxide moiety in TEMPO-Glc_4 bound to CBD_{N1} and CBD_{N2} were calculated using the measured r_{eff} as distance restraints for a simple minimization routine in X-PLOR (Brünger,

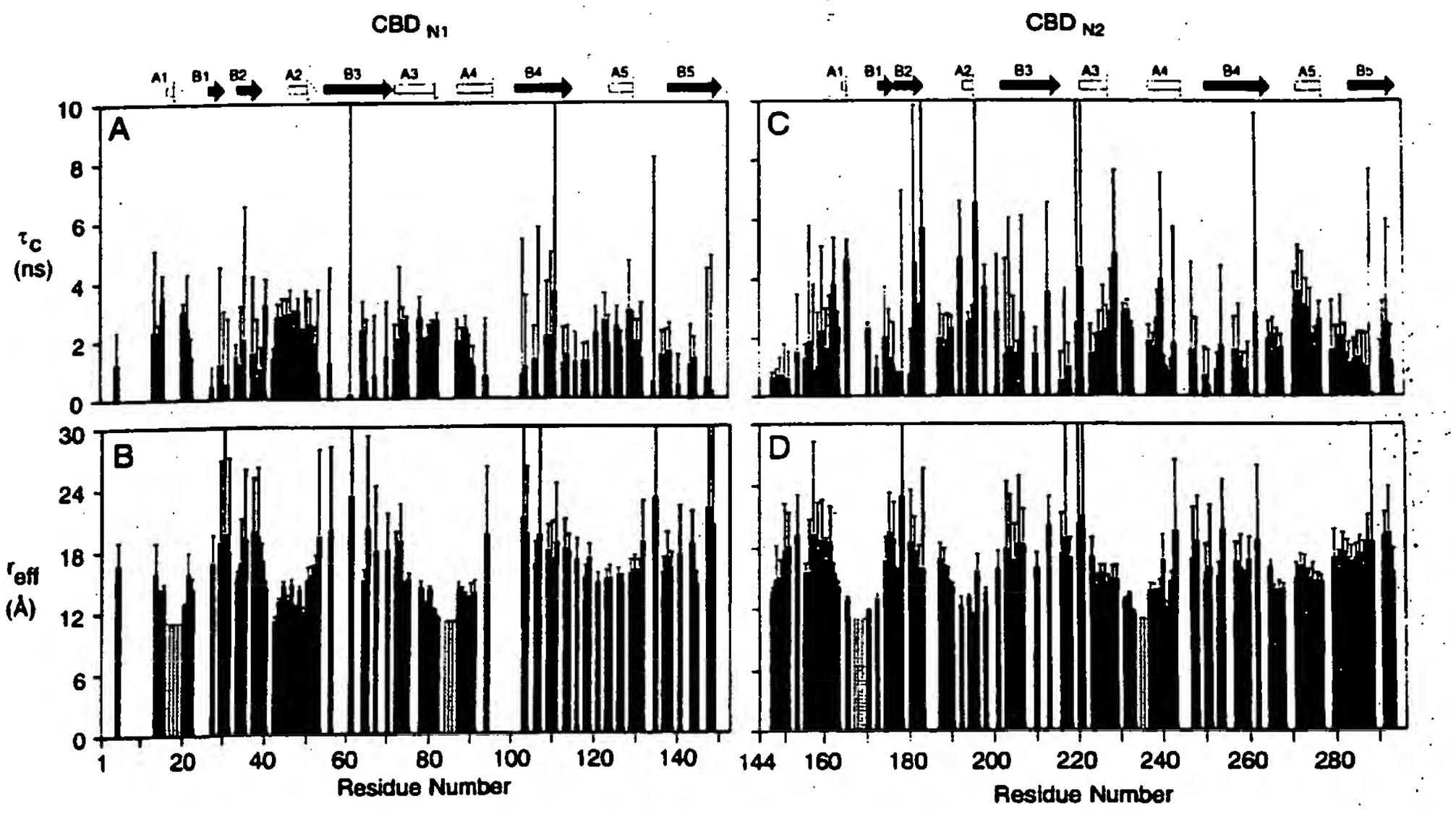


Figure 6. The correlation times, τ_c, and effective electron-proton distances, r_{eff}, calculated for (a), (b) CBD_{N1} and (c), (d) CBD_{N2} bound to TEMPO-Glc_4. Residues whose signals disappear completely due to severe linebroadening are shown with a shaded bar at an upper bounds of $r_{eff} = 11$ Å. The locations of the β-strands forming sheets A and B in the two proteins are indicated by the open and filled arrows, respectively.

1992). To account for multiple binding orientations, we used an ensemble-averaging protocol in which $r_{eff} = [1/n\ \Sigma r^{-6}]^{-1/6}$ where r is the electron-proton distance for each of n conformations, assumed to be equally populated (Bonvin & Brünger, 1996). A total of 60 minimization runs were performed using CBD_{N1} or CBD_{N2} in a system with $n = 1$, 2, or 3 nitroxide groups, modeled as "atoms" of mass 30 amu and a van der Waals radii of 2.5 Å. When only one nitroxide group is utilized, the "atom" is always located in the middle of the binding cleft, approximately above β-strand A5. However, this solution poorly satisfies the experimental restraints requiring the spin label to be closer to amide protons on the edges of the binding cleft than to those near its center. When an ensemble of two nitroxide atoms is considered, we find that in 31/60 and 60/60 cases for CBD_{N1} and CBD_{N2}, respectively, one atom is localized to each edge of the binding cleft (Figure 5). In the remaining 29/60 cases for CBD_{N1}, one atom is found near the middle of the binding cleft and the other distant from the protein. The lowest energies clearly occur in the former situation, with the nitroxide group near β-strands A1 and A4. The difference between the results obtained for the two proteins is attributed to the exact choice of the bounds used for the distance restraints. When an ensemble of three nitroxide atoms is generated, 24/60 and 42/60 cases for CBD_{N1} and CBD_{N2}, respectively, have one atom at each edge of the binding cleft and the third distant from the protein. In the remaining 36/60 and 18/60 cases, the nitroxide groups occupy a variety of positions, close to and distant from the proteins.

Based on these calculations, we conclude that the simplest explanation for the observed patterns of paramagnetic relaxation enhancement is that TEMPO-Glc_4 binds both CBD_{N1} and CBD_{N2} in two possible orientations such that the nitroxide moiety lies at either end of the binding cleft (Figure 5). In support of these results, the separation of ~23 Å between the positions calculated for the nitroxide atoms matches approximately the length of a TEMPO-Glc_4 molecule.

Discussion

Binding affinity for spin-labeled cellooligosaccharides

CBD_{N1} and CBD_{N2} bind TEMPO-Glc_3 and TEMPO-Glc_4 with affinities that are approximately equal to those measured previously for the unmodified sugars. Although the six-membered ring of TEMPO resembles in some ways that of a glucopyranosyl unit, it differs due to the presence of methyl substituents and the absence of equatorial hydroxyl groups. As discussed by Tomme *et al.* (1996a), calorimetric studies have revealed that the binding of soluble cellooligosaccharides to CBD_{N1} is driven by enthalpically favorable interactions such as hydrogen bonding and van der Waals contacts. We therefore conclude that the TEMPO group does not interact significantly with CBD_{N1} or CBD_{N2} and is positioned in or near the binding cleft solely by virtue of its covalent bonding to the cellooligosaccharides. This statement is supported by the observation that the chemical shift changes accompanying the binding of the modified and unmodified cellooligosaccharides to the two proteins are similar. We also conclude that the presence of an anomeric hydroxyl group does not contribute to the binding of cellooligosaccharides by CBD_{N1} or CBD_{N2}. This is not unexpected given that cellulose, the natural ligand for these proteins, exists in a highly polymeric form.

Multiple binding orientations

Qualitative and quantitative analyses of the effects of TEMPO-Glc_3 and TEMPO-Glc_4 on the relaxation properties of CBD_{N1} and CBD_{N2} clearly demonstrate that the labeled cellooligosaccharides are bound in at least two orientations. These orientations position the nitroxide moiety at either edge of the binding cleft of each CBD, such that amide protons in β-strands A1 and A4, as well as in the loop between A3 and A4, are perturbed by the spin label to a greater extent that those in the central strand A5 (Figure 5). Since the CBDs form 1:1 complexes with the cellooligosaccharides, these orientations are mutually exclusive such that in any individual CBD complex, the sugar is bound in one direction or the other. Given that the TEMPO-labeled sugars and the corresponding unmodified cellooligosaccharides bind to the CenC CBDs with similar affinities and produce similar chemical shift perturbations, we conclude Glc_3, Glc_4, and by inference, amorphous cellulose and other soluble β-glucans, are also bound in multiple orientations.

Binding of the unmodified and TEMPO-labeled cellooligosaccharides occurs on an NMR time-scale of fast exchange as evident by the progressive change in the chemical shifts of CBD_{N1} and CBD_{N2} upon the addition of these sugars. Therefore, the chemical shift of an amide proton at any point in the titration reflects a population-weighted average of its chemical shifts in the free and all-bound forms of the CBD. As a result, the equilibrium association constants determined by NMR and calorimetric methods represent apparent or macroscopic K_a values that are the summation of the microscopic binding constants for each possible orientation of the sugar-protein complex (Wyman & Gill, 1990).

An important question to ask is whether a preference for one binding orientation exists. During the titrations of CBD_{N1} and CBD_{N2} with the nitroxide-labeled sugars, it was noted that the signals from residues on β-strand A1 disappeared at lower levels of saturation than did those on A4 (Johnson, 1998). This could indicate that the binding of TEMPO-Glc_4 in an orientation with the nitroxide moiety near β-strand A1 is slightly favored, or simply that the nitroxide is positioned closer to the

amides in this strand relative to those in A4. If we assume that the TEMPO-labeled sugars bind in fast exchange between two possible orientations, then the measured r_{eff} is given by:

$$r_{eff} = \left(\frac{f_{b1}}{r_1^6} + \frac{f_{b2}}{r_2^6}\right)^{-1/6}$$

Here r_1 and r_2 are the distances from an amide proton to the two positions of the free electron, and f_{b1} and f_{b2} are the fractions of total protein bound by the ligand in each orientation. Using the positions of the nitroxide atoms shown in Figure 5, we calculate that varying f_{b1} from 0.1 to 0.9 does not change the predicted r_{eff} for amide protons near the edges of the CBD_{N1} and CBD_{N2} binding clefts by more ±20%. In other words, within experimental error, the effects of any small population differences between the spin-labeled ligands bound in one or the other orientation cannot be distinguished due to the sixth-power dependence of the relaxation enhancement on the proton-electron separation. This is readily seen by considering a case for which $r_1 = 10$ Å and $r_2 = 20$ Å. The calculated r_{eff} is 10.2, 11.2, and 14.4 Å for f_{b1} of 0.9, 0.5, and 0.1, respectively. A similar conclusion was made by Bonvin & Brünger (1995, 1996) when they demonstrated that NOE-derived distances are not sufficient to determine the fractional occupancies of the individual conformations of a protein that exhibits conformational heterogeneity. Because of this limitation, we conservatively estimate that the relative association constants for the binding of TEMPO-Glc_4 to CBD_{N1} or CBD_{N2} in the two possible orientations are within a factor of approximately five- to tenfold. Note also that if $r_1 = r_2$ and $f_{b1} = f_{b2}$, then saturation of the CBD at 65% ($f_{b1} = 0.325$) and 80% ($f_{b1} = 0.4$) leads to an r_{eff} that is overestimated by only ~7% and 4%, respectively.

Extrapolating from these studies, it is reasonable to suggest that TEMPO-Glc_3 and TEMPO-Glc_4 may be bound in multiple conformations that have one or the other of these two general orientations, yet also differing in the exact position (or register) of the sugar with the β-sheet clefts of CBD_{N1} and CBD_{N2}. Evidence for this hypothesis stems from two observations. First, the longer r_{eff} values (>15 Å) measured in this study, were consistently shorter than the electron-amide proton distances predicted from the calculated positions of the nitroxide atoms shown in Figure 5. Bearing in mind the significant errors associated with the measurement of these distances (see below), this could reflect a small population of ligand shifted to position the nitroxide near the center of the binding clefts of the CBDs and thus transiently close to amide protons in strand A5 and β-sheet B. Second, as noted previously (Johnson *et al.*, 1996a), the $^1H^N$ and ^{15}N chemical shifts of CBD_{N1} and CBD_{N2} are perturbed similarly upon binding of Glc_3, Glc_4, Glc_5, and Glc_6. This suggests that the four cellooligosaccharides interact structurally with the CBDs in the same manner, possibly by sliding within their binding clefts.

Structural symmetry can explain the multiple binding orientations

An explanation for the binding of cellooligosaccharides by CBD_{N1} and CBD_{N2} in multiple orientations is provided by inspection of the structures of both the ligands and the protein molecules. Hydrogen bond formation between the oligosaccharide and the protein provides an enthalpic driving force for binding (Tomme *et al.*, 1996a), and thus the locations of polar atoms within the molecules should be considered. As shown in Figure 7, the hydroxyl and hydroxymethyl groups of cellotetraose are approximately symmetrically disposed such that comparable hydrogen bonds could be formed to the CBDs whether the reducing end of the cellooligosaccharide is located closer to β-strand A1 or A4. In addition, as illustrated in Figure 8, the polar, non-polar, and aromatic sidechains implicated in cellooligosaccharide recognition (Johnson *et al.*, 1996a,b) are approximately symmetrically positioned about the centers of the binding clefts of CBD_{N1} and CBD_{N2}. This provides

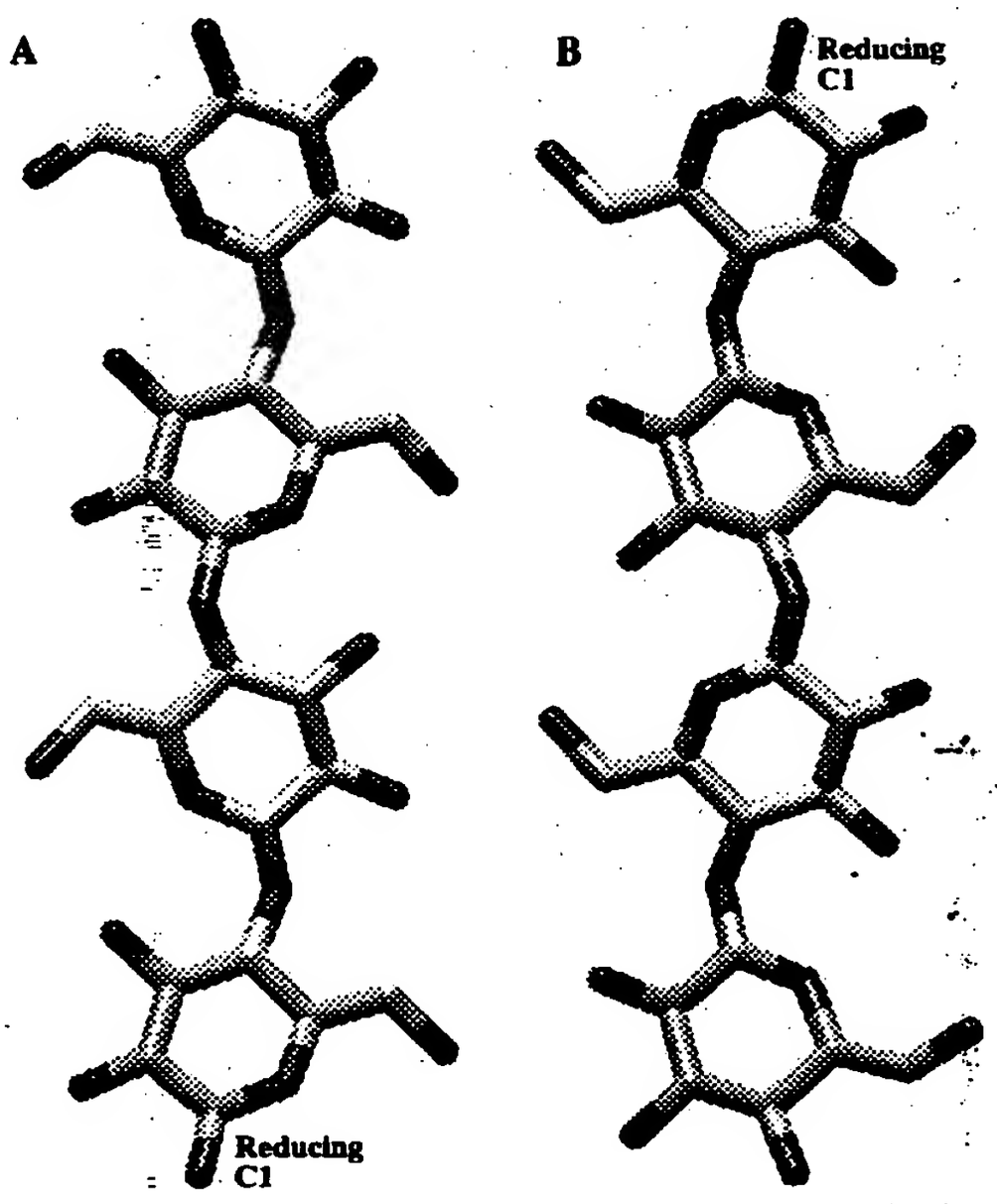


Figure 7. Cellotetraose has approximate symmetry such that its hydroxyl and hydroxymethyl groups would occupy similar positions when the sugar chain is bound in either orientation by CBD_{N1} and CBD_{N2}. (a), (b) Cellotetraose (Geßler *et al.*, 1994) with its reducing end positioned at the bottom or top of the Figure, respectively. The molecules are related by successive 180° rotations about axes perpendicular and parallel to the long axis of the cellooligosaccharide.

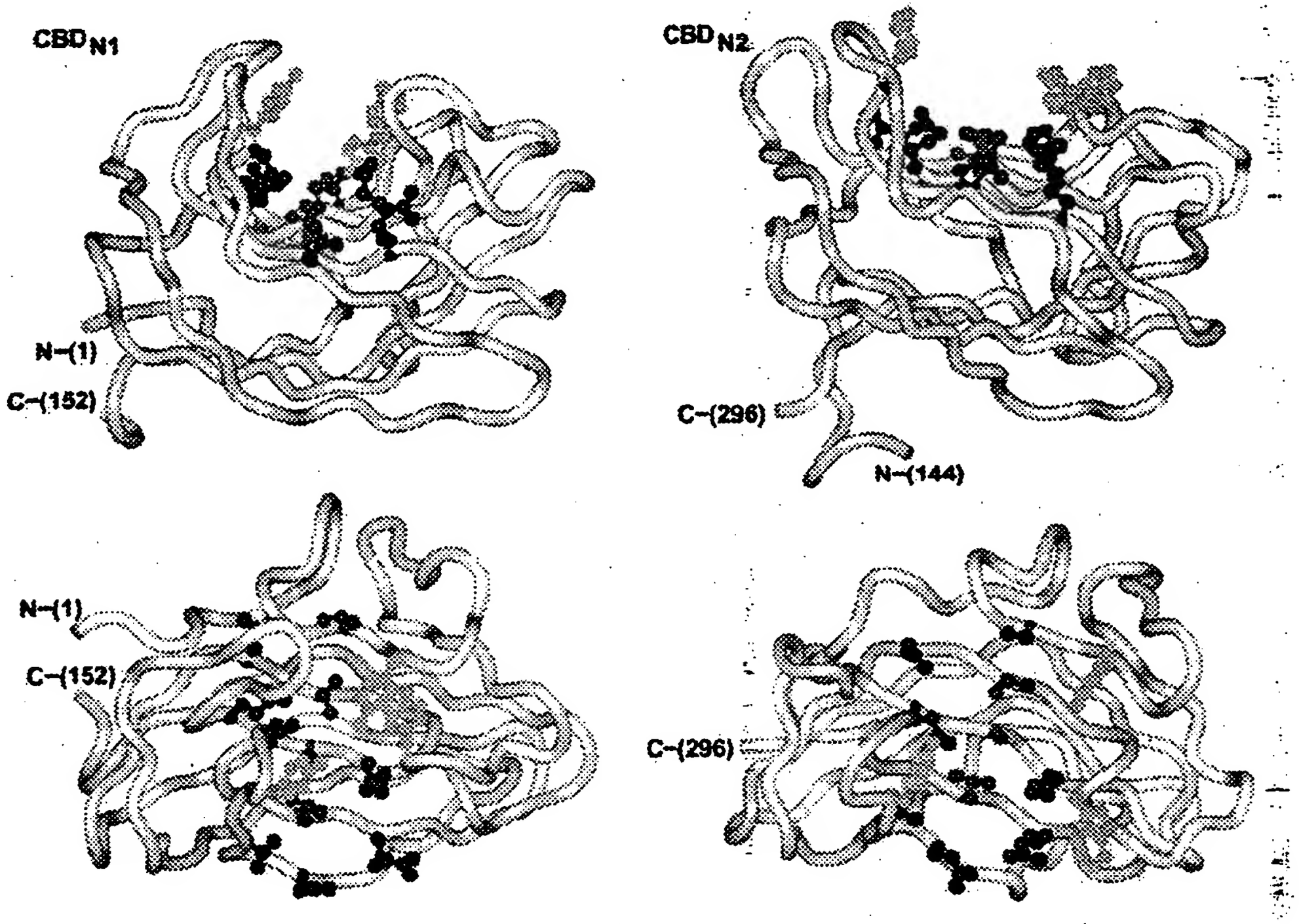


Figure 8. The residues implicated in cellulose recognition by the CenC CBDs are positioned approximately symmetrically about the centers of their binding clefts. This may facilitate the binding of a cellooligosaccharide chain in either orientation. Shown are ribbon diagrams of CBD_{N1} (left) and CBD_{N2} (right) positioned for the reader to look across (top) and down onto (bottom) the cellooligosaccharide binding cleft that is formed by β-sheet A in each protein. The side-chains of selected residues within the cellulose binding clefts are color coded as blue, polar; red, hydrophobic; and yellow, aromatic.

further opportunity for the complex formation of the cellooligosaccharides in multiple orientations, while maintaining similar patterns of hydrogen bonds to the equatorial hydroxyl groups and van der Waals contacts to the glucopyranose rings.

It is unlikely that all the oxygen atoms in cellotetraose and the polar atoms in the binding clefts of CBD_{N1} or CBD_{N2} are involved simultaneously in intermolecular hydrogen bonding. The association of cellotetraose with CBD_{N1} involves a ΔG° of -4.9 kcal mol^{-1} at 35 °C (Tomme *et al.*, 1996a). If the sole contribution to this binding free energy is assumed to be hydrogen bonding involving the 14 sugar hydroxyl groups, each would account for only -0.35 kcal mol^{-1}. Thermodynamic studies of several protein-carbohydrate complexes have demonstrated that the presence of some hydroxyl groups is essential for binding, whereas others are unnecessary or even a hindrance. The net contributions of individual hydroxyl groups to binding generally range from -0.5 kcal mol^{-1} to -2 kcal mol^{-1}, yet can be as high -6 kcal mol^{-1} (Lemieux, 1996; Quiocho, 1993; Street *et al.*, 1986). Based on these values, CBD_{N1} likely forms between 2.5 and ten hydrogen bonds to cellotetraose. Given the large number of potential hydrogen-bonding sites on this sugar and on the side-chains lining the binding clefts of CBD_{N1} and CBD_{N2}, it is reasonable to suggest that cellooligosaccharides, as well as amorphous cellulose, could bind these proteins in many possible conformations with similar free energies. These conformations include opposite orientations of the sugar chain across the β-sheet faces of the CBDs, as well as possible multiple registers within their binding clefts.

Further insights into the binding promiscuity of the CenC CBDs are also gained from dynamic studies of the proteins and protein-sugar complexes using ^{15}N and ^{2}H relaxation measurements. As will be reported elsewhere, the backbone amide protons and methyl-containing side-chains of CBD_{N1} show motional disorder on a nano- to picosecond time-scale that is not restricted upon binding cellopentaose. This internal mobility or flexibility may also allow the cellooligosaccharides to bind in multiple, rapidly interconverting conformations.

Use of spin labels for obtaining ligand-protein distance restraints

Nitroxide-labeled cellooligosaccharides clearly provide a powerful method for characterizing protein-carbohydrate interactions. Qualitatively, the effects of the free radical on the NMR spectra of CBD_{N1} and CBD_{N2} yielded immediate evidence for the binding of TEMPO-Glc_3 and Glc_4 in multiple orientations. Although it was also possible to calculate the effective positions of the nitroxide moiety bound in a minimum of two possible orientations on the CBDs from a quantitative analysis of the measured paramagnetic relaxation enhancement, the accuracy of these models is limited for numerous reasons. These include: (i) complications associated with multiple conformations and partial saturation, as discussed above; (ii) the uncharacterized conformational mobility of the bound TEMPO group; (iii) inaccuracies in the NMR-derived structures of CBD_{N1} and CBD_{N2} calculated without inclusion of the bound Glc_4 or Glc_5, respectively (Johnson *et al.*, 1996b); (iv) errors, which tend to lead to an underestimation of longer values of r_{eff}, arising from difficulties in measuring small values of ΔR_1 and ΔR_2, and from the possibility of relaxation enhancement due to nonspecific binding or to collisions between CBD-TEMPO-Glc_4 complexes in solution; and (v) the lack of distance restraints with $r_{eff} < 11$ Å resulting from the efficiency of paramagnetic relaxation. The latter problem could be addressed by studying CBD_{N1} and CBD_{N2} partially titrated with small quantities of TEMPO-Glc_4 or with mixtures of the labeled sugar in its oxidized and reduced forms in order to identify the amides closest to the bound nitroxide.

To understand in detail the mechanism of cellulose binding by the two Family IV CBDs, we are currently exploiting glycosynthases to prepare selectively ^{13}C-labeled cellooligosaccharides. These compounds will be utilized for isotope-edited and filtered experiments aimed at resolving and assigning intermolecular NOEs between the CBDs and the bound sugars. Using a combination of short-range distance restraints derived from interproton NOE interactions and long-range restraints from electron-proton paramagnetic relaxation enhancements, we hope to define more accurate structural models of the ensembles of the complexes of CBD_{N1} and CBD_{N2} with cellooligosaccharides.

Biological implications

Studies of numerous cellulases and hemicellulases indicate that CBDs facilitate hydrolysis by passively increasing local enzyme concentrations on the surface of cellulosic and hemicellulosic substrates and, possibly, by actively disrupting noncovalent interactions between the chains of these polysaccharides (Tomme *et al.*, 1995). The specificity of CBDs for the various allomorphs of cellulose (e.g. crystalline or amorphous) will also serve to target these enzymes to distinct regions of this complex substrate.

The results presented here demonstrate clearly that in isolation, CBD_{N1} and CBD_{N2} each bind soluble cellooligosaccharides in multiple orientations. This immediately prompts the question as to the role played by these CBDs within their native context. CenC is a 1069 residue β-1,4-glucanase, composed of CBD_{N1} and CBD_{N2} arranged in tandem at its N terminus, a central family 9 catalytic domain, and two additional domains of unknown function at its C terminus (Coutinho *et al.*, 1992, 1991). The enzyme cleaves a variety of cellulosic substrates and, based on an analysis of its hydrolysis of carboxymethyl-cellulose, is semi-processive with both endo- and exoglucanase activities(Tomme *et al.*, 1996b). Although the structures of the isolated family IV binding domains (CBD_{N1} and CBD_{N2}) and the homologous Family 9 catalytic domains from *Clostridium thermocellum* celD (Juy *et al.*, 1992) and *Thermomonospora fusca* E4 (Sakon *et al.*, 1997) have been determined, the spatial arrangement of these modules within native *C. fimi* CenC remains to be defined.

CBD_{N1} and CBD_{N2} are juxtaposed without an intervening linker. Given that the affinity of the each domain for cellotetraose is equivalent whether joined or in isolation (unpublished), it is likely that CBD_{N1N2} can also bind oligomeric sugars in multiple orientations. This theory is reasonable, as the binding clefts of the two homologous CBDs both span approximately five glycosyl units and thus each domain in CBD_{N1N2} can independently accommodate a single cellooligosaccharide molecule. Preliminary studies indicate that the affinity of CBD_{N1N2} for phosphoric acid swollen cellulose is only approximately 2 fold greater than that of CBD_{N1} for this polymeric substrate (Tomme *et al.*, 1996a). This implies that, when joined in tandem, these CBDs bind amorphous cellulose in an additive, rather than co-operative, manner. A simple interpretation of this result is that the two domains comprising CBD_{N1N2} are structurally constrained, due to the lack of a flexible linker, such that they cannot bind simultaneously to adjacent regions of a single polymer chain. We therefore speculate that the tandem CBDs anchor CenC to its natural substrate by bridging chains of amorphous cellulose without a strong preference for their orientations. Note however, as discussed above, we can only estimate that the relative association constants for the binding of TEMPO-Glc_4 to CBD_{N1} or CBD_{N2} in the two possible orientations are within a factor of approximately five- to tenfold. Thus in combination, CBD_{N1N2} may in fact exhibit a significant overall preference for binding strands of amorphous cellulose positioned to match the yet unknown orientation of the two domains within this tandem CBD. Resolution of this important question hinges upon the determination of the structure of CBD_{N1N2}, as well as further characterization of naturally occurring amorphous cellulose (Bayer *et al.*, 1998).

CBD_{N1N2} is connected to the catalytic domain of CenC by a short proline-rich polypeptide linker sequence (Coutinho *et al.*, 1991, 1992). In contrast to the isolated CBDs, it is certain that this catalytic domain has a definite orientational requirement for the binding and hydrolysis of cellulosic substrates. Although the exact boundaries and thus length of the linker sequence is unknown, the observation that it is highly sensitive to proteolytic cleavage (Tomme *et al.*, 1996b) suggests that the linker affords at least some degree of flexibility between the binding and catalytic domains of CenC. This may allow the cleavage of properly oriented chains of amorphous cellulose other than those to which the CBDs are bound. Processivity would result from the diffusional sliding of both the binding and catalytic domains along the substrate cellulose chains. Enzymatic studies of CenC with and without CBD_{N1N2}, as well as structural characterization of the catalytic domain, are clearly necessary to understand the role of these Family IV CBDs in modulating the activity of this *C. fimi* β-1,4-glucanase.

Although over 180 putative CBDs in 13 families have been identified, little is known about the detailed mechanism of cellulose and hemicellulose binding by these protein domains. Studies of several family 2 CBDs, such as those from *C. fimi* Cex, have demonstrated that binding to crystalline cellulose is mediated by conserved aromatic residues aligned along a flat face of the molecule (Xu *et al.*, 1995). Calorimetric measurements have proven that binding is entropically driven(Creagh *et al.*, 1996), indicative of hydrophobic stacking of the aromatic side-chains on the glucose rings. Hydrodynamic studies have also revealed that CBD_{Cex} diffuses readily across the surface of crystalline cellulose (Jervis *et al.*, 1997). We therefore speculate that CBD_{Cex}, as well as members of other CBD families, also binds in multiple orientations to cellulose. In contrast, based on the crystal structure of a fragment of the endo/exocellulase E4 containing a family 3c CBD and a family 9 catalytic domain, Sakon *et al.* (1997) have hypothesized that the binding domain aids in the activity of this enzyme by helping to feed a cellulose chain, bound in a defined orientation, into the active site of the catalytic domain. This distinctive behaviour may result from the intimate structural association of the two domains in the *T. fusca* endo/exocellulase.

In summary, we have shown that CBD_{N1} and CBD_{N2} from *C. fimi* CenC bind TEMPO-labeled cellooligosaccharides in multiple orientations. This promiscuous mode of sugar recognition can be attributed to the approximate symmetry of the hydrogen bonding groups on cellulose and on the side-chains lining the binding clefts of these two CBDs. Further structural studies of both native CenC and its substrate, amorphous cellulose, are necessary to understand the mechanisms by which this and other cellulolytic enzymes function in the efficient degradation of biomass.

Materials and Methods

Synthesis of TEMPO-labeled cellooligosaccharides

Details of the synthesis of TEMPO-Glc_3 and TEMPO-Glc_4 will be provided elsewhere. Ionspray mass spectroscopy of the final products yielded 658 (M+), 659 (M + 1) and 660 (M + 2) for TEMPO-Glc3 (calculated for $C_{27}H_{48}NO_{17}$: 658.7) and 820 (M+), 821 (M + 1), and 822 (M + 2) for TEMPO-Glc_4 (calculated for $C_{33}H_{58}NO_{22}$: 820.8). Useful ^{1}H-NMR data could not be obtained for these products due to the presence of the spin label. Structural characterization was therefore achieved by reduction of the nitroxide group to a hydroxylamine and acetylation of the entire molecule. The spectral data for these derivatives of TEMPO-Glc_3 and TEMPO-Glc_4, respectively, are provided below.

2,2,6,6-Tetramethylpiperidin-1-acetoxyl-4-yl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl-(1→4)-2,3,6-tri-O-acetyl-β-D-glucopyranosyl-(1→4)-2,3,6-tri-O-acetyl-β-D-glucopyranoside

^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.14 (dd, 1 H, $J_{3,4}$ 9.6 Hz, H-3), 5.13-5.09 (m, 2 H, H-3′ H-3″), 5.05 (dd, 1 H, $J_{4'',5''}$ 9.5 Hz, H-4″), 4.89 (dd, 1 H, $J_{2'',3''}$ 8.8 Hz, $J_{2'',1''}$ 8.1 Hz, H-2″), 4.85 (dd, 1 H, $J_{2',3'}$ 9.5 Hz, $J_{2',1'}$, 8.7 Hz, H-2′), 4.83 (dd, 1 H. $J_{2,3}$ 9.6, Hz $J_{2,1}$ 7.3 Hz, H-2), 4.60-4.47 (m, 4 H, H-1 H-1′ H-1″ H6a), 4.42 (dd, 1 H. $J_{6a'',6b''}$ 11.7 Hz, $J_{6a'',5''}$ 1.5 Hz, H-6a′), 4.34 (dd, 1 H, $J_{6a'',6b''}$ 12.4 Hz, $J_{6a'',5''}$ 4.4 Hz, H-6a″), 4.14 (dd, 1 H, $J_{6b',5'}$ 4.5 Hz, H-6b′), 4.07 (dd, 1 H, $J_{6b,5}$ 5.9 Hz, H-6b), 4.04 (dd, 1 H, $J_{6b'',5''}$ 1.6 Hz, H-6b″), 3.93 (m, 1 H, H-1‴), 3.76 (dd, 1 H, $J_{4',5'}$ 9.4 Hz, H-4′), 3.71 (dd, 1 H, $J_{4,5}$ 9.5 Hz, H-4), 3.65 (m, 1 H, H-5″), 3.63-3.58 (m, 2 H, H-5 H-5′), 2.14-1.98 (33 H, 11 × Ac), 1.83-1.38 (m, 4 H, H-2‴ H-3‴), 1.13 (6 H, CH_3), 1.10 (6 H, CH_3). MS (Ionspray) 1123 (M + 1); calculated for $C_{49}H_{71}NO_{28}$ (1122.1).

2,2,6,6-Tetramethylpiperidin-1-acetoxyl-4-yl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl-(1→4)- 2,3,6-tri-O-acetyl-β-D-glucopyranosyl-(1→4)- 2,3,6-tri-O-acetyl-β-D-glucopyranosyl -(1→4)- 2,3,6-tri-O-acetyl-β-D-glucopyranoside

^{1}H NMR (400 MHz, $CDCl_3$) δ: 5.15 (dd, 1 H, $J_{3,4}$ 9.5 Hz, H-3), 5.12-5.08 (m, 3 H, H-3′ H-3″ H-3‴), 5.04 (dd, 1 H, $J_{4''',5'''}$ 9.5 Hz, H-4‴), 4.89 (dd, 1 H, $J_{2''',3'''}$ 8.9 Hz, $J_{2''',1'''}$ 8.0 Hz, H-2‴), 4.86- 4.81 (m, 3 H, H-2 H-2′ H-2″), 4.60-4.47 (m, 5 H, H-1 H-1′ H-1″ H-1‴ H6a), 4.60-4.32 (m, 3 H, H-6a′ H-6a″ H-6a‴), 4.15-4.00 (m, 4 H, H-6b H-6b′ H-6b″ H-6b‴) 3.95 (m, 1 H, H-1⁗), 3.81-3.68 (m, 3 H, H-4 H-4″ H-4⁗), 3.66 (m, 1 H, H-5‴), 3.62-3.57 (m, 3 H, H-5 H-5′ H-5″), 2.15-1.96 (42 H, 14 x Ac), 1.85-1.37 (m, 4 H, H-2⁗ H-3⁗), 1.15-1.10 (12 H, 4 × CH_3). Ionspray MS 1411 (M + 1); calculated for $C_{61}H_{87}NO_{36}$ (1410.34).

Protein samples

The ^{15}N-labeled CBD_{N1} (residues 1-152 of CenC) and CBD_{N2} (residues 146-296) were produced by expression of the plasmids pTugN1 and pTugN2 in *Escherichia coli* JM101 cells grown in minimum media containing 1 g/l of $^{15}NH_4Cl$ and 1 g/l of 99% ^{15}N-labeled Isogro (Isotec Inc.; Johnson *et al.*, 1996a). The secreted proteins were purified from the culture supernatant and the periplasmic fraction using both affinity chromatography on cellulose (Avicel) and ion-exchange chromatography, as

described (Johnson *et al.*, 1996a). Samples of CBD_{N1} and CBD_{N2} were exchanged into a final buffer of 50 mM sodium chloride, 50 mM sodium $^2H_3$-acetate, 3 to 5 mM $CaCl_2$, and 0.02% sodium azide in 10% ^2H_2O/90% H_2O at pH* 6.1 using a microsep concentration device (Filtron). CBD_{N1} binds calcium with an equilibrium association constant of $\sim 10^5$ M^{-1} at pH 6.0 and thus exists as a 1:1 CBD_{N1}-Ca^{+2} complex under these conditions (Johnson *et al.*, 1998). The metal ion stabilizes the folded structure of CBD_{N1} but does not alter its affinity for cellooligosaccharides. In contrast, CBD_{N2} does not bind calcium appreciably.

Titrations with TEMPO-labeled cellooligosaccharides

The binding of the TEMPO-labeled cellooligosaccharides to the CenC CBDs at pH* 6.1 and 35 °C was measured quantitatively using ^1H-^{15}N HSQC spectroscopy. Stock solutions of TEMPO-Glc_3 and TEMPO-Glc_4 were prepared by weight in the above buffer and added in small aliquots to the protein samples. The titration of CBD_{N1} with TEMPO-Glc_3 was carried out by progressively adding the labeled sugar to 0.32 mM protein up to a final ligand-to-protein ratio of 22:1. The titrations of 0.65 mM CBD_{N1} and 0.48 mM CBD_{N2} with TEMPO-Glc_4 were taken to final corrected ligand-to-protein ratios of 2.2:1 and 1.4:1, respectively. Protein concentrations were measured using $\varepsilon_{280} = 21370$ M^{-1} cm^{-1} and 20500 M^{-1} cm^{-1} for CBD_{N1} and CBD_{N2}, respectively (Johnson *et al.*, 1996a). Equilibrium association binding constants were determined by non-linear least-squares fitting of the $^1H^N$ and ^{15}N chemical shifts of numerous amide protons *versus* cellooligosaccharide concentration to the Langmuir isotherm describing the binding of one ligand molecule to a single protein site(Johnson *et al.*, 1996a). These amide protons, which include Gly7, Gly15, Val34, Gly44, Thr65, Gln80, Asn81, Thr87 and Gly130 for CBD_{N1} and Val197, Tyr198, Gly231, Tyr234 and Ala237 for CBD_{N2}, exhibit fast exchange between the free and bound states on the chemical shift time-scale and are detectable, albeit with diminishing intensity, over the course of the entire titration. In the cases of the TEMPO-Glc_4 titrations, good fits of the experimental data points to the binding isotherms were obtained only if scaling factors of ~1.3 (for CBD_{N2}) or 1.8 (for CBD_{N1}) were introduced to correct for apparent errors in measuring the weight of the sugar (~1 mg) during the preparation of stock titrant solutions. The reported binding constants and errors are the averages and standard deviations of the individual values measured for each $^1H^N$ and ^{15}N nuclei.

NMR spectroscopy

Spectra were acquired on a Varian Unity 500 MHz spectrometer at 35 °C. Each point of the titration of the ^{15}N-labeled proteins with TEMPO-Glc_3 or Glc_4 was monitored by the acquisition of a sensitivity-enhanced gradient ^1H-^{15}N HSQC spectrum (Kay *et al.*, 1992). Upon formation of the sugar-protein complexes, $^1H^N$ T_1 values were measured using a sensitivity-enhanced gradient ^1H-^{15}N HSQC sequence as a read-out of a non-selective inversion-recovery experiment recorded with delays of $t = 0$, 0.1, 0.2, 0.4, 0.8 and 2 seconds. The $t = 0.1$ spectrum was repeated to help estimate experimental error. A non-sensitivity-enhanced gradient ^1H-^{15}N HSQC spectrum was then recorded to obtain the spin-spin relaxation enhancements, ΔR_2, of the amide $^1H^N$ nuclei. After completion of the data collection with either CBD_{N1} or CBD_{N2} bound to the paramagnetic cellooligosaccharide, the nitroxide functionality was reduced to the diamagnetic hydroxylamine by the addition of two molar equivalents of solid L-ascorbic acid (Sigma). The pH* of the sample was re-adjusted to 6.1 and the spectra necessary to measure ΔR_1 and ΔR_2 were recorded. All spectra were acquired as 1024 × 96 complex points with spectral widths of 6500 and 1450 Hz in the ^1H and ^{15}N dimensions, respectively. Selective water flip back pulses were utilized to ensure minimum perturbation of the water magnetization (Zhang *et al.*, 1994; Grzesiek & Bax, 1993).

Calculation of ΔR_1 and ΔR_2

NMR spectra were analyzed using a combination of Felix (Biosym Technologies) and NMRPipe (Delaglio *et al.*, 1995). Data for the ^1H T_1 series were processed with mild Lorentzian-to-Gaussian apodization. The relative peak volumes, V_t, at each inversion-recovery delay t were fit to the function $V_t = V_0 \exp(-t/T_1)$, as described by Farrow *et al.* (1994). V_0 is the volume at time $t = 0$, and errors in the measured T_1 lifetimes were estimated using Monte Carlo simulations. The paramagnetic enhancement of each amide proton spin-lattice relaxation rate, ΔR_1, was calculated as:

$$\Delta R_1 = \left(\frac{1}{T_1}\right)_{ox} - \left(\frac{1}{T_1}\right)_{red} \quad (1)$$

Ox and red indicate the data collected with the labeled cellooligosaccharides in the nitroxide and hydroxylamine forms, respectively.

The non-sensitivity enhanced ^1H-^{15}N HSQC spectra were processed without apodization in the proton dimension and with a mild exponential linebroadening window in the nitrogen dimension. Two methods were used to measure individual values of ΔR_2. In the first, the ratio of the peak volumes recorded for an amide in the presence of the nitroxide (V_{ox}) *versus* the hydroxylamine-labeled (V_{red}) cellooligosaccharide is given by:

$$\frac{V_{ox}}{V_{red}} = \frac{e^{-t}/T_{2_{ox}}}{e^{-t}/T_{2_{red}}} \quad (2)$$

where $t = 10.1$ ms is the total time during the INEPT and reverse-INEPT components of the HSQC pulse sequence in which $^1H^N$ magnetization is in the transverse plane and thus subject to paramagnetic relaxation. A non-gradient enhanced HSQC was utilized to avoid possible complications due to the use of two reverse-INEPT sequences. By measuring total peak volumes, and not heights, the effect of differential T_2 relaxation during the detection periods of the spectra recorded in the presence of the oxidized and reduced sugars can be neglected. Peak volumes were determined by direct integration over a manually adjusted "footprint" and not by fitting of an idealized peak shape. Equation (2) can be rearranged to obtain $\Delta R_{2,Vol}$:

$$\Delta R_{2,Vol} = \left(\frac{1}{T_2}\right)_{ox} - \left(\frac{1}{T_2}\right)_{red} = \left(\frac{1}{t}\right) \ln\left(\frac{V_{red}}{V_{ox}}\right) \quad (3)$$

In the second method, $\Delta R_{2,LW}$ were obtained from the change in the linewidths of the ^1H-^{15}N cross-peaks in the proton dimension:

$$\Delta R_{2,LW} = \left(\frac{1}{T_2}\right)_{ox} - \left(\frac{1}{T_2}\right)_{red} = \pi(LW_{ox} - LW_{red}) \quad (4)$$

where LW_{ox} and LW_{red} are the full half-height line-widths in the presence of the nitroxide and hydroxylamine-labeled cellooligosaccharides, respectively. The line-widths were measured by non-linear least-squares fitting of vectors, extracted in the proton dimension through a given cross-peak, to the equation for a single Lorentzian peak split by a $^3J_{HN\text{-}H\alpha}$ coupling using the program PLOTDATA (TRIUMPH, UBC, Vancouver, Canada). The coupling constant was either treated as an independent variable or fixed at the value measured previously (Johnson *et al.*, 1996b) using the HNHA or HMQC-J experiments (Bax *et al.*, 1994; Kay & Bax, 1990). Significant errors in the measurement of $\Delta R_{2,LW}$ arise if this coupling is ignored. The effects of the free electron on the relaxation rates of the amide nitrogens are neglected due to the small magnetogyric ratio of the ^{15}N nucleus relative to that of ^1H.

Calculation of r_{eff}

The magnetic interaction of an unpaired electron and a proton is described by the modified Solomon-Bloembergen equations (Solonon & Bloembergen, 1956).The enhancement of the longitudinal (ΔR_1) and transverse (ΔR_2) relaxation rates of a proton due to the spin label are given by the equations (Kosen, 1989; Gellespie & Shortle, 1997a):

$$\Delta R_1 = \Delta\left(\frac{1}{T_1}\right) = \frac{2K}{r^6}\left(\frac{3\tau_c}{1+\omega_H^2\tau_c^2}\right) \quad (5)$$

$$\Delta R_2 = \Delta\left(\frac{1}{T_2}\right) = \frac{K}{r^6}\left(4\tau_c + \frac{3\tau_c}{(1+\omega_H^2\tau_c^2)}\right) \quad (6)$$

where K is the constant 1.23×10^{-32} cm^6 s^{-2} for a nitroxide radical, r is the distance between the electron and proton, τ_c is the correlation time for the fluctuation of the electronic-nuclear dipole-dipole interaction, and ω_H is the Larmor frequency of the proton. These equations are based on the assumptions that the vector between the electron and proton is free to undergo isotropic rotational diffusion, and that the distance r is constant. The value of τ_c for each individual amide can be determined directly from the ratio of ΔR_2 to ΔR_1 by combining equations (1) and (2) to obtain:

$$\tau_c = \left(\frac{6\left(\frac{\Delta R_2}{\Delta R_1}\right) - 7}{4\omega_H^2}\right)^{1/2} \quad (7)$$

With a value of τ_c in hand, the distance between the proton and electron, r, is calculated by rearranging equation (2):

$$r^6 = \frac{K}{\Delta R_2}\left(4\tau_c + \frac{3\tau_c}{(1+\omega_H^2\tau_c^2)}\right) \quad (8)$$

Individual values of τ_c, rather than a global average, were utilized to reflect directly the experimentally measured relaxation enhancements for each amide. For an ensemble of interconverting protein-ligand complexes, the measured distances will reflect the weighted-average of the individual electron-proton distances and thus are denoted as r_{eff}.

Standard deviations for ΔR_1, ΔR_2, τ_c and r_{eff} were calculated by conventional error propagation (Bevington & Robinson, 1992).

Structure modeling

The positions of the nitroxide moiety in TEMPO-Glc_4 bound to CBD_{N1} or CBD_{N2} were estimated using a simple restrained minimization protocol provided with X-PLOR v.3.8 (Brünger, 1992). The only attractive force was the soft-square potential of the distance restraint energy term (E_{NOE}) and the only repulsive force was that corresponding to van der Waals interactions (E_{vdw}). The bounds on the effective electron-proton distances, r_{eff}, were defined for three cases. First, for eight amide protons in CBD_{N1} and eight in CBD_{N2} for which ΔR_2 values could not be obtained due to severe paramagnetic line-broadening (thus $r_{eff} < 11$ Å) , the distance bounds were set to the range of 1-11 Å. Second, for 27 amide protons in CBD_{N1} and 26 in CBD_{N2} for which the measured r_{eff} were 11 to 15 Å, the distances were restrained to $r_{eff} \pm 3$ Å. Third, for 54 amides in CBD_{N1} and 59 in CBD_{N2} with $r_{eff} > 15$ Å, the distance bounds were set from 13 Å to an arbitrarily high value of 45 Å. Although distances between 11 and 23 Å were measured for CBD_{N1} and CBD_{N2} (Figure 6), it was found that those >15 Å were consistently underestimated in light of the dimensions of these proteins. Therefore, r_{eff} values greater than 15 Å were used only to derive lower bounds to restrain the nitroxide atoms from approaching the corresponding amide protons in the two protein structures.

To account for multiple positions of TEMPO-Glc_4 bound to CBD_{N1} or CBD_{N2}, ensemble-averaging (Bonvin & Brünger, 1996) was applied to coordinates systems containing one, two, or three nitroxides, defined as single atoms with a mass of 30 amu and a van der Waal radius of 2.5 Å. The co-ordinates of the protein (pdb file 1ULO for CBD_{N1} and a low-energy structure for CBD_{N2} (unpublished)) were held fixed, and each ensemble of nitroxide groups was minimized 60 times starting from random positions. The quality of the calculated structures were estimated by the NOE energy penalty due to distance restraint violations.

Acknowledgements

We are grateful to Lewis Kay, Douglas Kilburn, Peter Tomme, and Antony Warren for insightful discussions. This work was funded by the Government of Canada's Network of Centres of Excellence program supported by the Medical Research Council of Canada and the Natural Sciences and Engineering Research Council through PENCE Inc. P.E.J. and E.B. contributed equally to this work.

References

Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). Cellulose, cellulases, and cellulosomes. *Curr. Opin. Struct. Biol.* **8**, 548-557.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. & Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79-105.

Bevington, P. R. & Robinson, D. K. (1992). *Data Reduction and Error Analysis for the Physical Sciences*, 2nd edit., McGraw-Hill Inc., Toronto.

Bonvin, A. M. J. J. & Brünger, A. T. (1995). Conformational variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.* **250**, 80-93.

Bonvin, A. M. J. J. & Brünger, A. T. (1996). Do NMR distances contain enough information to assess the relative populations of multi-conformer structures? *J. Biomol. NMR*, **7**, 72-76.

Brünger, A. (1992). *X-PLOR: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. (1996). *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, Toronto.

Coutinho, J. B., Moser, B., Kilburn, D. G., Warren, R. A. J. & Miller, R. C., Jr. (1991). Nucleotide sequence of the endoglucanase C gene (CenC) of *Cellulomonas fimi*, its high-level expression in *Escherichia coli*. and characterization of its products. *Mol. Microbiol.* **5**, 1221-1233.

Coutinho, J. B., Gilkes, N. R., Warren, R. A. J., Kilburn, D. G. & Miller, R. C., Jr. (1992). The binding of *Cellulomonas fimi* endoglucanase C (CenC) to cellulose and sephadex is mediated by the N-terminal repeats. *Mol. Microbiol.* **6**, 1243-1252.

Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. & Haynes, C. A. (1996). Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl Acad. Sci. USA*, **93**, 12229-12234.

Creagh, A. L., Koska, J., Johnson, P. E., Tomme, P., Joshi, M. D., McIntosh, L. P., Kilburn, D. G. & Haynes, C. A. (1998). Stability and oligosaccharide binding of the N1 cellulose-binding domain of *Cellulomonas fimi* Endoglucanase CenC. *Biochemistry*, **37**, 3529-3537.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Foreman-Kay, J. D. & Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, **33**, 5984-6003.

Geßler, K., Krauß, N., Steinr, T., Betzel, C., Sandmann, C. & Saenger, W. (1994). Crystal structure of β-D-cellotetraose hemihydrate with implications for the structure of cellulose II. *Science*, **266**, 1027-1029.

Gellespie, J. R. & Shortle, D. (1997a). Characterization of long-range structure in the danatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158-169.

Gellespie, J. R. & Shortle, D. (1997b). Characterization of long-range structure in the danatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**, 170-184.

Gnewuch, T. & Sosnovsky, G. (1986). Spin-labeled carbohydrates. *Chem. Rev.* **86**, 203-238.

Grzesiek, S. & Bax, A. (1993). The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593-12594.

Jervis, E. J., Haynes, C. A. & Kilburn, D. G. (1997). Surface diffusion of cellulases and their isolated binding domains on cellulose. *J. Biol. Chem.* **272**, 24016-24048.

Johnson, P. E. (1998). Structural and dynamic analysis of oligosaccharide binding by CBD_{N1}. PhD thesis, University of British Columbia, Vancouver, Canada.

Johnson, P. E., Tomme, P., Joshi, M. D. & McIntosh, L. P. (1996a). Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 2. NMR and ultraviolet absorption spectroscopy. *Biochemistry*, **35**, 13895-13906.

Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G. & McIntosh, L. P. (1996b). Structure of the N-terminal cellolose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance. *Biochemistry*, **35**, 14381-14394.

Johnson, P. E., Creagh, L. A., Brun, E., Joe, K., Tomme, P., Haynes, C. A. & McIntosh, L. P. (1998). Calcium binding by the N-terminal cellulose-binding domains from *Cellulomonas fimi* β-1,4-glucanase CenC. *Biochemistry*, **37**, 12772-12781.

Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Béguin, P. & Aubert, J.-P. (1992). Three-dimensional structure of a thermostable bacterial cellulase. *Nature*, **357**, 89-91.

Kay, L. & Bax, A. (1990). New methods for the measuring of NH-CαH coupling constants in ^{15}N-labelled proteins. *J. Magn. Reson.* **86**, 110-126.

Kay, L., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665.

Kleerekoper, Q., Howarth, J. W., Guo, X., Solaro, R. J. & Rosevear, P. R. (1995). Cardiac troponin I induced conformational changes in cardiac troponin C as monitored by NMR using site-directed spin and isotope labeling. *Biochemistry*, **34**, 13343-13352.

Kosen, P. A. (1989). Spin labeling of proteins. *Methods Enzymol.* **177**, 86-121.

Lemieux, R. U. (1996). How water provides the impetus for molecular recognition in aqueous solution. *Acc. Chem. Res.* **29**, 373-380.

Mackenzie, L. F., Wang, Q., Warren, R. A. J. & Withers, S. G. (1998). Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* **120**, 5583-5584.

Otting, G. & Wüthrich, K. (1990). Heteronuclear filters in two-dimensional [^1H,^1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Quart. Rev. Biophys.* **23**, 39-96.

Plessas, N. R. & Goldstein, I. J. (1981). Synthesis of α- and β-glycosides containing spin labels as probes for studies of carbohydrate-protein interactions. *Carbohydrate Res.* **89**, 211-220.

Quiocho, F. (1989). Protein-carbohydrate interactions: basic molecular features. *Pure Appl. Chem.* **61**, 1293-1306.

Quiocho, F. A. (1993). Probing the atomic interactions between proteins and carbohydrates. *Biochem. Soc. Trans.* **21**, 442-448.

Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). Structure and mechanism of endo/exocellulase E4 from Thermonospora fusca. *Nature Struct. Biol.* **4**, 810-818.

Solomon, L. & Bloembergen, N. (1956). Nuclear magnetic interactions in the HF molecule. *J. Chem. Phys.* **25**, 261-266.

Street, I. P., Armstrong, C. R. & Withers, S. G. (1986). Hydrogen bonding and specificity. Fluorodeoxy sugars as probes of hydrogen bonding in the glyco-

gen phosphorylase/glucose complex. *Biochemistry*, 25, 6021-6027.

Tomme, P., Warren, R. A. J., Miller, R. C., Jr, Kilburn, D. G. & Gilkes, N. R. (1995). Cellulose-binding domains: classification and properties. In *Enzymatic Degradation of Insoluble Polysaccharides* (Saddler, J. N. & Penner, M., eds), pp. 142-161, American Chemical Society, Washington, DC.

Tomme, P., Creagh, A. L., Kilburn, D. G. & Haynes, C. A. (1996a). Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 1. Binding specificity and calorimetric analysis. *Biochemistry*, 35, 13885-13894.

Tomme, P., Kwan, E., Gilkes, N. R., Kilburn, D. G. & Warren, R. A. J. (1996b). Characterization of CenC, an enzyme from *Cellulomonas fimi* with both endo- and exoglucanase activities. *J. Bacteriol.* 178, 4216-4223.

Vyas, N. K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* 1, 732-740.

Wyman, J. & Gill, S. J. (1990). *Binding and Linkage. Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley.

Xu, G.-Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E. & Harvey, T. S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry*, 34, 6993-7009.

Yu, L., Meadows, R. P., Wagner, R. & Fesik, S. W. (1994). NMR studies of the FK506 binding protein bound to a spin-labeled ascomycin analog. *J. Magn. Reson.* 104, 77-80.

Zhang, O., Kay, L. E., Olivier, J. P. & Foreman-Kay, J. D. (1994). Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domains of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR*, 4, 845-858.

Zhao, Q., Abeygunawardana, C. & Mildevan, A. S. (1997). NMR studies of the secondary structure in solution and the steroid binding site of delta5-3-ketosteroid isomerase in complexes with diamagnetic and paramagnetic steroids. *Biochemistry*, 36, 3458-3472.

Edited by P. E. Wright

(Received 8 December 1998; received in revised form 8 February 1999; accepted 11 February 1999)

http://www.academicpress.com/jmb

Supplementary material comprising one Figure is available from JMB Online

Reprinted from *J. Mol. Biol.* (1999) **293**, 107–115

JMB

Residual Dipolar Coupling Derived Orientational Constraints on Ligand Geometry in a 53 kDa Protein-Ligand Complex

P. J. Bolon, H. M. Al-Hashimi and J. H. Prestegard

Article No. jmbi.1999.3133 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 293, 107–115

JMB

Residual Dipolar Coupling Derived Orientational Constraints on Ligand Geometry in a 53 kDa Protein-Ligand Complex

P. J. Bolon, H. M. Al-Hashimi and J. H. Prestegard*

Complex Carbohydrate Research Center, University of Georgia Athens, GA 30602, USA

**Corresponding author*

The geometric relationships between ligands and the functional groups that bind ligands in soluble ligand-protein complexes have traditionally been deduced from distance constraints between pairs of NMR active nuclei spanning the ligand-protein interface. Frequently, the steep inverse distance dependence of the nuclear Overhauser effect (NOE), from which the distance constraints are derived, makes identification of sufficient numbers of constraints difficult. In these cases the ability to supplement NOE-derived information with distance-independent angular information can be very important. Here, the observation of residual dipolar couplings from α-methyl mannose bound to mannose binding-protein in a dilute liquid crystalline medium has allowed the determination of a bound ligand's average orientation. The 3-fold rotational symmetry of mannose-binding protein defines its orientational tensor and obviates the need to determine experimentally the protein's average orientation. Through superimposition of ligand and protein orientational tensors we describe the binding geometry of α-methyl mannose bound to mannose-binding protein. This new method is of general applicability to the study of ligands bound to proteins, and it is of particular interest when neither X-ray crystallography nor NOE techniques can provide sufficient information to describe binding geometries.

Keywords: dipolar couplings; NMR; structure determination; carbohydrate interactions; mannose-binding protein

Introduction

We illustrate here a new approach for the determination of ligand geometry in protein binding sites that relies on orientational constraints derived from residual dipolar couplings observed in NMR spectra of partially ordered macromolecular complexes (Bothner-By, 1995; Prestegard, 1998; Prestegard *et al.*, 1998; Tjandra & Bax, 1997; Tolman *et al.*, 1995). The determination of the bound geometry of protein ligands is an important issue for improving our understanding of how structure relates to biological function, and also for building a basis for rational drug design. Many useful approaches to this determination have been developed in the past, including X-ray diffraction of crystalline protein-ligand complexes and NMR-based transfer NOE studies of complexes in solution (Ni, 1994). All the approaches are limited in certain ways. In the case of X-ray crystallography, complexes must be crystallizable; in the case of transferred nuclear Overhauser effect (NOE) studies, ligands must be in rapid exchange. Transfer NOE studies, while giving sound information on ligand geometry, also suffer from a lack of information on the nature of ligand protein contacts. Information on these contacts can be provided by more conventional NOE studies using both protein and ligand resonances in some cases, but this information is often difficult to obtain because it requires assignment of protein as well as ligand resonances, and because it presumes sufficiently close approach of protein and ligand protons to allow efficient magnetization transfer ($r < 5$ Å).

Cases where determination of bound ligand geometry have been particularly difficult include those involving protein-oligosaccharide interactions.

Abbreviations used: NOE, nuclear Overhauser effect; MBP, mannose-binding protein; CRD, carbohydrate recognition domain; AMM, α-methyl mannoside; HSQC, heteronuclear single quantum coherence.

E-mail address of the corresponding author: jpresteg@ccrc.uga.edu

0022-2836/99/410107–9 $30.00/0

These cases encompass an important area in that protein-oligosaccharide interactions mediate a variety of processes essential to cell function and survival. To name a few, sialyl Lewis-X oligosaccharide-selectin interactions mediate the recruitment of leucocytes to sites of injury. The interaction of choleratoxin with the oligosaccharide headgroup of ganglioside GM_1 mediates internalization of the toxic subunit, and polylactosamine-galectin interactions are hypothesized to be involved in the metastasis of malignant cells (Lasky, 1992). Characterization by traditional NMR methods of the oligosaccharide-protein interactions that mediate these processes is difficult because the hydrogen-bonding networks involving hydroxyl protons on the sugars are often part of the interface between protein and oligosaccharide. The hydroxyl protons are then the ones in closest proximity to protein protons, but these exchange rapidly with protons in bulk water making their NMR resonances hard to observe. In these difficult cases, residual dipolar data can offer a valuable alternative, where the data can in principle constrain both bound ligand geometry and ligand orientation relative to the protein binding site.

Our particular target in the studies presented here is mannose-binding protein-A (MBP). MBP is a protein important in innate immune response. It binds sugars, including α-mannosides, that are frequently exposed on the surface of invading pathogens, and subsequently triggers a complement cascade (Ikeda *et al.*, 1987). Improper activation *via* MBP has also been suggested as a factor in inflammatory disease (Dwek, 1996). Crystal structures of several forms of MBP are available. The form we study here is a 149 residue truncated version that includes a calcium-containing carbohydrate recognition domain (CRD) and a short α-helical stem that initiates the formation of a homo-trimer by making a left-handed coiled-coil with helical stems from two other MBP monomers. The 3-fold symmetry axis in the resulting homo-trimeric structure is important for the interpretation of our results. The binding of several ligands to MBP and mutant MBPs that mimic selectin proteins has also been studied, both because of interest in MBP's ligand binding properties and those of proteins which show homology to the MBP CRD, such as the selectins. Crystallography has been the primary contributor in studies of MBP trimers, with structures existing for bound galactoside and galactosamine, as well as sialyl Lewis-X and related ligands (Chang *et al.*, 1994; Iobst *et al.*, 1994; Kolatkar AR, 1998; Ng & Weis, 1997; Weis & Drickamer, 1994). Additionally, there are structures of a shorter fragment of MBP which makes a homo-dimer; one of these includes α-methyl mannoside in the carbohydrate binding site (Ng *et al.*, 1996). The results of the above crystallography studies provide a useful basis for evaluating data from our new approach.

The new approach that we present is based on the observation of residual dipolar coupling of directly bonded ^1H -^{13}C pairs in α-methyl mannoside (AMM) when the mannoside is bound to MBP in a field-oriented aqueous liquid crystal. The use of residual dipolar coupling in the characterization of the structures of proteins and other macromolecules in solution has received a good deal of attention recently (Clore *et al.*, 1998a; Tjandra & Bax, 1997) . When molecules are partially ordered in a magnetic field interaction, vectors connecting magnetic nuclei in the molecules depart from a complete isotropic sampling of all directions in space. This gives rise to a contribution to splitting of NMR resonances that depends on the magnetic properties of the interacting nuclei, the internuclear separation (r) and the angle (θ) that the vector makes with the magnetic field. When the nuclei are directly bonded as in a ^1H-^{15}N or ^1H-^{13}C pairs, the $1/r^3$ distance-dependence of the interaction can be regarded as being determined by the bond length and the primary variable becomes the orientation of the bond relative to the magnetic field as characterized by the angle θ. The functional dependence of the dipolar couplings is $(1/2(3\cos^2\theta - 1))$, where the parentheses denote an average over an orientation distribution, and the resulting interaction appears as an addition to the normal one bond scalar couplings. These through space interactions are always present; in fact, they are the basis of the common NOE interaction. However, the angular function above is averaged to zero when space is isotropically sampled and direct splittings of resonances only appear in partially oriented media. Fortunately, several means of producing field-ordered states are now available, including inherent orientation due to the large anisotropies in the magnetic susceptibilities of some molecules (Kung *et al.*, 1995; Tolman *et al.*, 1995), orientation due to the interaction of molecules with lipid bicelles that form field-oriented liquid crystals (Tjandra & Bax, 1997) and interaction with filamentous bacteriophage that also form field-oriented liquid crystals (Clore *et al.*, 1998c; Hansen *et al.*, 1998).

The interpretation of residual dipolar data has taken two routes, incorporation of individual bond constraints as penalty functions in simulated annealing protocols for molecular structure determination (Clore *et al.*, 1998a,b), and extraction of order tensors that directly describe the direction and level of ordering forces from the point of view of a coordinate frame fixed in a rigid molecular fragment (Losonczi *et al.*, 1999; Losonczi & Prestegard, 1998b; Saupe, 1968). The latter approach which arose from substantial previous work in the liquid crystal field, proves to be particularly valuable for the problem at hand. An order tensor is a 3×3 matrix with elements $\langle 1/2(3\cos\theta_i\cos\theta_j - \delta_{ij})\rangle$ written in terms of direction cosines that relate Cartesian axes of an arbitrarily chosen molecular axis system to the magnetic field direction. Because the matrix is traceless and symmetric, there are only five independent elements. An order matrix in an arbitrarily chosen frame of a molecular fragment is difficult to inter-

pret, but if diagonalized, the five independent variables turn into a principal order parameter, an asymmetry parameter and three Euler angles relating the diagonal, or principal frame, to the original molecular frame. If fragments of interest experience a common source of order, as they would if they were parts of a rigid ligand-protein complex, the directions and levels of the orienting force should appear the same from the point of view of each fragment when a model for the fragments is assembled with proper fragment orientations. This provides a test for the validity of proposed models for the orientation of a ligand (one fragment) in a protein binding site (another fragment).

The key to using this test is having an approximately rigid fragment with enough potential dipolar coupled pairs to determine the five elements of an order tensor. For a ^{15}N-labeled protein, pairs of ^{1}H-^{15}N spins are clearly abundant and much of the protein backbone can be considered rigid. In oligosaccharides, individual pyranose rings frequently prefer a $^{4}C_1$ configuration which can be considered to have approximately rigid geometry and ^{1}H-^{13}C pairs are abundant. In α-methyl mannoside there are five directly bonded ^{1}H-^{13}C pairs on the $^{4}C_1$ ring, two of which point in two unique directions and three of which depart slightly from a third direction. These pairs can provide the data for an order tensor determination. There are a number of ways of accurately determining the splittings of resonances from these pairs (Tolman & Prestegard, 1996) but the easiest is simply to collect ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectra, without decoupling in one of the frequency domains. This is the approach we use here. We used uniformly ^{13}C-labeled α-methyl mannoside in order to improve sensitivity.

Results and Discussion

Measurement of residual dipolar couplings in AMM complexed to MBP

The ^{1}H-^{13}C HSQC NMR spectra of ^{13}C-enriched AMM in the presence of an equimolar amount of MBP dissolved in a dilute bicelle medium are shown in Figures 1(a) and (b) at temperatures of 25 °C and 39 °C, respectively. The data were acquired in the absence of ^{13}C decoupling during acquisition, and the one bond ^{13}C-^{1}H couplings are shown as frequency domain splittings in the proton dimension. The magnitudes of these couplings and the corresponding errors extracted using a Bayesian parameter estimation method (Andrec & Prestegard, 1998) are shown on the spectra. As one would expect for the case of isotropic tumbling at 25 °C, all the couplings in Figure 1(a) are similar and fall in the range of typical $^{1}J_{CH}$ scalar coupling constants for non-anomeric sites in carbohydrates (140-150 Hz). On the other hand, couplings measured at 39 °C are more widely spread because of the presence of a residual dipolar contribution. Residual dipolar contributions for individual C-H bond vectors calculated from the differences in couplings measured at 25 °C (isotropic) and 39 °C (aligned) are shown in Table 1 (D_{obs}).

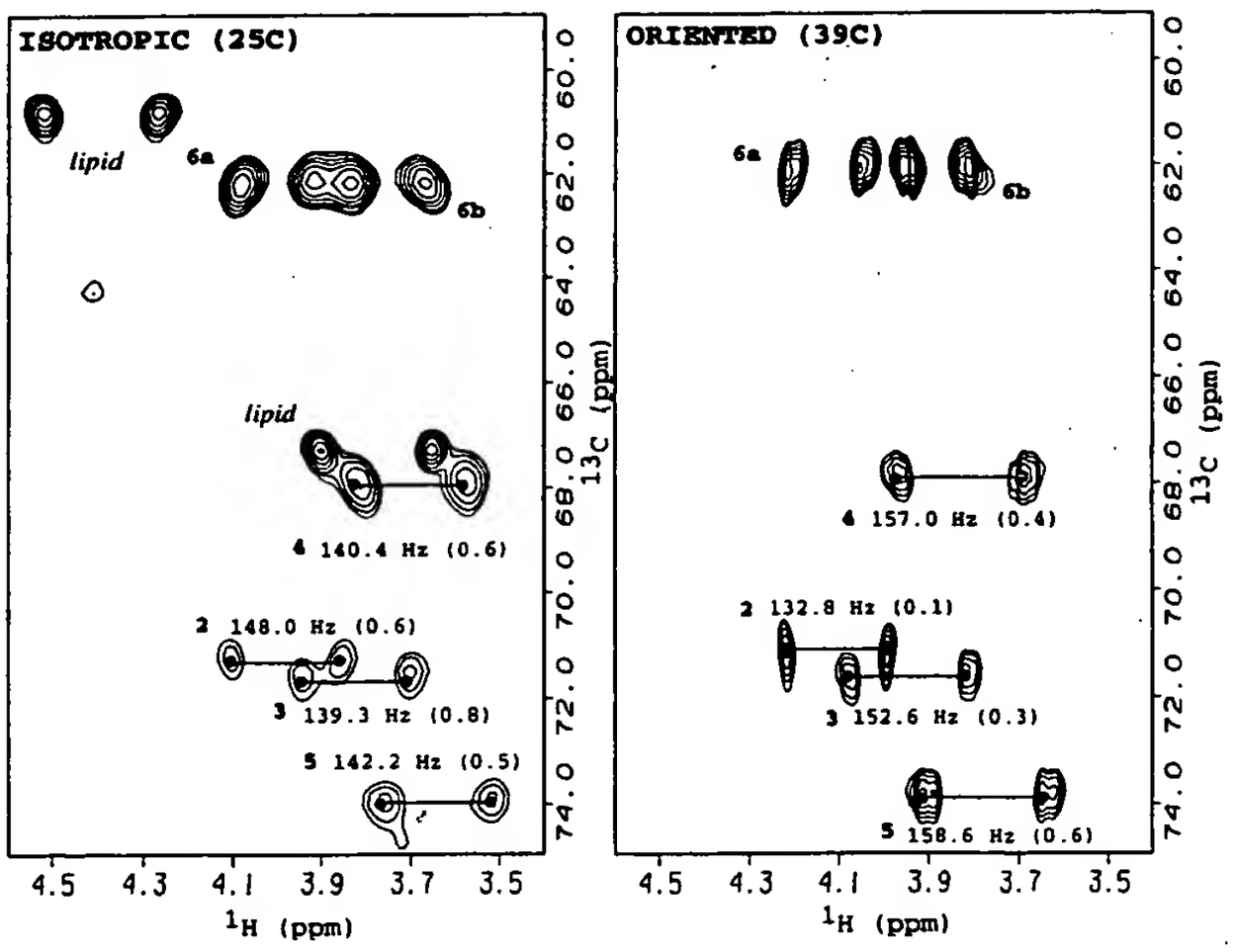


Figure 1. (a) Section of a proton-coupled HSQC spectrum of isotropic AMM in MBP with selected couplings annotated. (b) Section of a proton-coupled HSQC spectrum of oriented AMM in MBP with selected couplings annotated. The anomeric regions of the spectra are not shown but show splittings of 169.4(±0.5) Hz for isotropic condition and 158.6(±0.2) Hz for the oriented case.

Table 1. Residual dipolar couplings for bound and free AMM

Data Set	C_1-H_1 (Hz)	C_2-H_2 (Hz)	C_3-H_3 (Hz)	C_4-H_4 (Hz)	C_5-H_5 (Hz)
AMM + MBP + bicelle (D_{obs})	−10.8±0.7	−15.2±0.7	13.3±1.2	16.6±1	16.4±1.1
AMM + bicelle (D_{free})	−3.5±0.7	−7.3±0.7	9.2±0.9	12.1±0.7	7.8±0.5
AMM-bound state (D_{bound})	−22.6±2.2	−28.7±2.2	20±3.5	23.9±2.9	30.5±3.0

While residual dipolar coupling contributions are indeed observed in the AMM/MBP complex dissolved in a bicelle medium, they need not originate fully from binding of AMM to a strongly aligned MBP molecule. We see only a single set of resonances, but for a weak binding ligand in fast exchange that spends a portion of its time bound to the protein and the remainder free in solution, this will always be the case. Thus, observed splittings will be a population weighted average of those induced in the free and bound states. Residual dipolar couplings have previously been observed in free carbohydrate molecules dissolved in bicelle media (Bolon & Prestegard, 1998; Kiddle & Homans, 1998; Rundlof *et al.*, 1998). While small, they can be significant. In order to quantify the contribution from the AMM-free state, the same experiments were repeated under identical conditions in the absence of MBP. The differences in measured couplings between 25 °C and 39 °C are also shown in Table 1 (D_{free}). Again, couplings and associated errors were extracted using a Bayesian parameter estimation method. It is noteworthy that the magnitudes of couplings for various C-H bond vectors are generally smaller for free AMM samples, but the reduction varies from site to site. The different pattern of splittings assures us that we are not looking at a simple scaling of bicelle-induced order and that MBP has a significant direct effect on both extent and direction of AMM alignment. This direct binding effect was confirmed in another set of experiments. Here, a molar ratio of 1:8 MBP:AMM was used, which would reduce the bound AMM state to approximately 8%. Indeed, in this case where indirect effects would dominate, the presence of MBP had little measurable effect on the observed residual dipolar couplings.

Interpretation of residual dipolar contributions of a ligand in a protein-bound state will, in general, require separation of the contribution from the free state using known binding properties. The observed residual dipolar couplings measured in the AMM/MBP complex (Table 1, D_{obs}) will be a population weighted average from the free (N_{free}) and bound (N_{bound}) states, such that:

$$D_{obs} = N_{free}D_{free} + N_{bound}D_{bound} \quad (1)$$

The fraction of AMM free and bound (N_{free} and N_{bound}) can be calculated from the dissociation constant and known protein-ligand concentrations. Using a previously determined binding constant of 1 mM (Sayers, 1998), approximately 40% of AMM is in fact bound to MBP leaving approximately 60% of AMM free in solution. Residual dipolar couplings originating from the bound state (D_{bound}) calculated from equation (1) are shown in Table 1. Since these dipolar contributions originate from MBP's partial alignment, they contain the desired orientational information on AMM relative to MBP. This information is best extracted using an order matrix analysis.

Order tensor calculation in AMM

Using AMBER-minimized structure coordinates for AMM (Woods *et al.*, 1995) and five residual dipolar couplings measured in the sugar ring in simple bicelle solution, the five elements of the order matrix were determined using a singular value decomposition approach (Losonczi *et al.*, 1999). The initial molecular coordinate frame for AMM was defined by placing the *y*-axis along the C1-H1 vector and the *z*-axis normal to the plane defined by H1-C1-OMe. Using this frame, the position of the principal order frame axes determined from residual dipolar couplings from free AMM (D_{free}, Table 1) is illustrated in Figure 2(a) using a Sauson-Flaumsteed projection (Bugayevskiy & Snyder, 1995). The alignment tensor is extremely asymmetric and the directions of all three principal axes of the alignment tensor are well defined, except that inversion of any axis is allowed. The asymmetry is quantitatively defined by an asymmetry parameter eta (η) that can be written in terms of order parameters S_{ij} ($\eta = (S_{yy} - S_{xx})/S_{zz}$). It is equal to 0.8(±0.13). Asymmetric behavior is not unusual and has, in fact, been observed in many systems dissolved in bicelle media. Due to the high level of asymmetry, the direction of highest order (S_{zz}) can alternate between two orthogonal orientations (Figure 2(a)) for which S_{zz} assumes values similar in magnitude but opposite sign ($S_{zz} = +0.0005$ or -0.0005). Note that one of the possibilities for the higher level of order is nearly along the *x*-axis of the initial molecular coordinate frame (the center of the Sauson-Flaumsteed plot). Hence, it is nearly in the plane of the H1-C1-OMe fragment and perpendicular to the C-H vector.

The five elements of the order tensor for AMM determined from couplings for the MBP-bound state (D_{bound}) were also determined using a singular value decomposition method, and the directions of the principal axes are depicted in Figure 2(b). While the position of the *z*-axis, the most strongly ordered axis, is well defined, the position of the S_{xx} and S_{yy} axes are ambiguous.

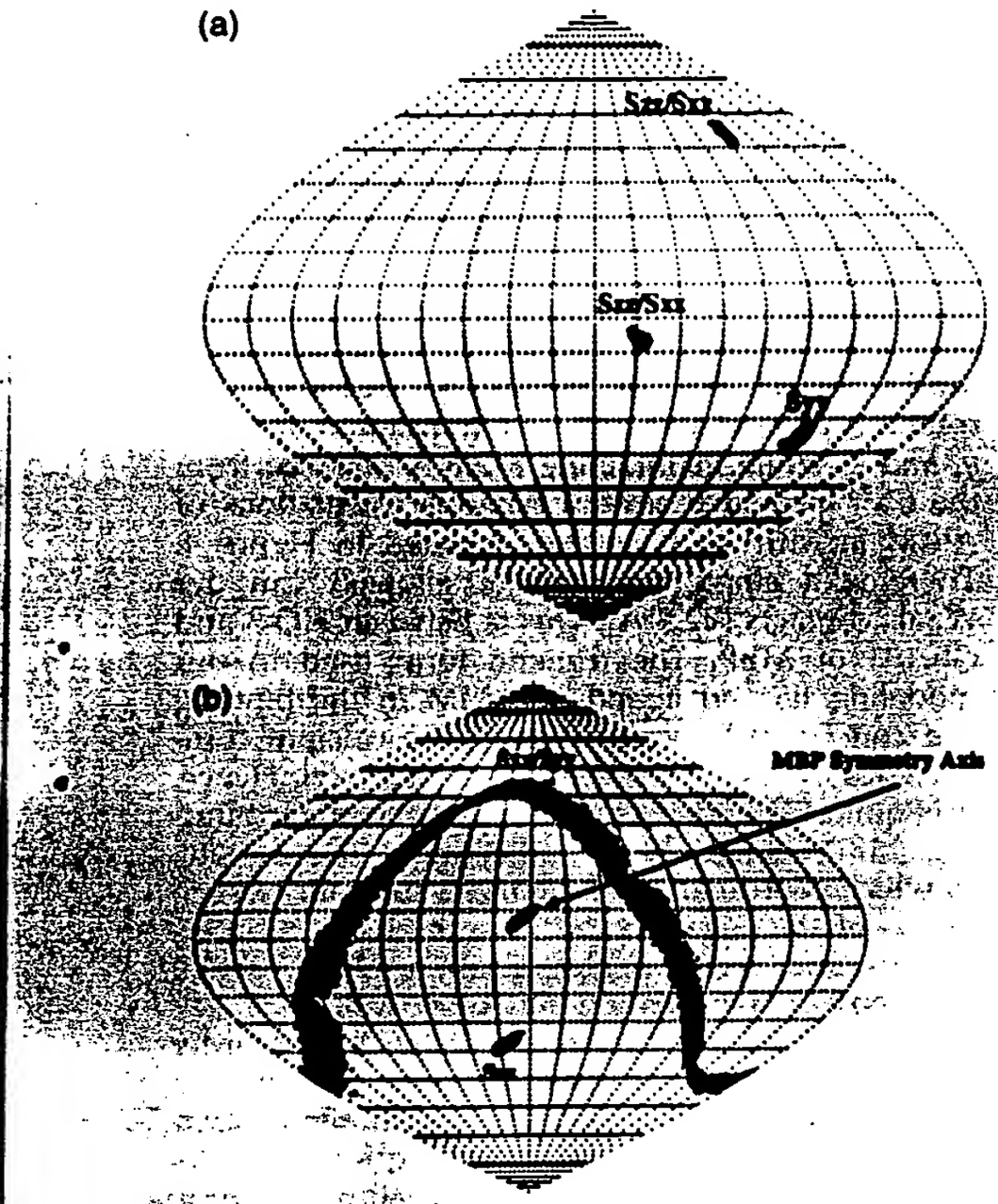


Figure 2. (a) Sauson-Flaumsteed projection of the directions of highest order for oriented AMM without MBP. (b) Sauson-Flaumsteed projection of the directions of highest order for oriented AMM in the presence of MBP, with a clear depiction of axial symmetry. The molecular coordinate frame for AMM was defined by placing the *y*-axis along the C1-H1 vector and the *z*-axis normal to the plane defined by H1-C1-OMe.

This is a characteristic of an axially symmetric order tensor. The asymmetry parameters confirms this axially symmetric nature of the alignment tensor with $\eta = 0.24(\pm 0.16)$. In addition, the level of order ($S_{zz} = -0.0012(\pm 0.00012)$) is approximately 2.5 times larger than that for AMM in free solution, and the direction of highest order is shifted 40° toward the molecular *z*-axis (out of the plane of to the H1-C1-OMe fragment).

The direction of alignment for the bound ligand is most useful when referenced to the alignment of MBP. We know that the direction of highest order (S_{zz}) depicted in Figure 2(b) should coincide with the direction of highest order for MBP itself. Rotating the molecular frame of the sugar to achieve coincidence then allows determination of the relative orientation of AMM bound to MBP. In general, determination of the alignment from a protein is possible using ^{1}H-^{15}N data as we have recently demonstrated for other systems (Fischer *et al.*, 1999). However, careful consideration of MBP's structure, and in particular, its symmetry, allows us to omit this independent interpretation and indeed explain the appearance of axial symmetry depicted in Figure 2(b).

MBP's alignment tensor: a solution in symmetry

The form of MBP used in these studies is a homo-trimer with a 3-fold symmetry axis and three equivalent sugar binding sites related by the same symmetry operation as depicted in Figure 3. The computed residual dipolar couplings originating from the AMM in the bound state (D_{bound}) will thus be an average over three orientations that are related by MBP's 3-fold symmetry axis. As has long been recognized and used in studies of small molecules (Saupe, 1968) this averaging has a very specific effect on order matrices determined from averaged couplings; regardless of the nature of MBP's interaction with the bicelle medium, AMM's experimentally determined direction of highest order must always point along the symmetry axis of MBP and the resulting order tensor must always be axially symmetric (Al-Hashimi *et al.*, 1999).

In the absence of any large conformational flexibility and mobility in the bound state, the order parameters determined from bound AMM ($S_{zz} = -0.0012(\pm 0.00012)$) will reflect MBP's order parameter. The fact that the principal order parameter is negative indicates that MBP's symmetry axis is, on average, perpendicular to the magnetic field and parallel with the normals of the bicelles in our

Figure 3. Depiction of AMM in fast exchange with the three equivalent CRDs of MBP. The 3-fold axial symmetry results in a direction of highest order along the rotor axis. Yellow spheres correspond to Ca^{2+}, black and red shperes to carbon and oxygen, respectively, of AMM, and MBP is represented by ribbon diagram.

medium (these order with their normals perpendicular to the magnetic field; Sanders *et al.*, 1994). This is as one would expect if interactions with the surfaces of bicelles dictated order of MBP. Axial symmetry should also propagate to the bound ligand and this is exactly what we observe in the order tensor determination using residual dipolar contributions from MBP bound AMM as shown in Figure 2(b). The direction of highest order, S_{zz}, in Figure 2(b) should now point along the symmetry axis of MBP. As we discuss below, the orientation of AMM deviates by about 40° from expected results using modeling studies and available crystal structures.

The orientation of AMM bound to MBP

No crystal structure of trimeric MBP-A with AMM in the binding site actually exists. However, AMM in the site of a dimer of the homologous MPB-C, missing the helical stem exists, and AMM in the trimer can be positioned by overlaying a carbohydrate-binding domain from the dimer on the three domains in the trimer structure (PDB entries 1KMB and 1RDL, rmsd 0.63-0.78 Å). The orientation of AMM, when positioned so that the experimentally determined symmetry axes coincide with that of MBP cannot, however, be made to overlap with the orientation of AMM modeled in this way. It is displaced by approximately 40°. There are several possible explanations for this observed deviation. First, our use of the trimer as a model may be in error. Instead, an equilibrium among monomer, dimer and trimer forms may exist in solution. This possibility is unlikely, since linewidths and diffusional properties of the construct used in previous studies are consistent with a trimer (Sayers, 1998). Moreover, the experimentally determined axially symmetric tensor is consistent with the expected 3-fold averaging. Second, the actual geometry of the trimer in solution may be different from what is depicted in the crystal structure. This is not out of the question; there are in fact X-ray structures of mutated MBPs in trimeric aggregates which show substantial differences in the relative placement of the CRDs, even though mutations are concentrated in loops at the extremities. For example, the helices near the trimer junction in PDB entries 1KMB and 1AFA are displaced from one another by more than a 1 Å when the stems are aligned. Similarly, the presence of internal motion in each protein domain or in AMM in the bound state can lead to deviations of average orientations. Third, it may be that alternate binding geometries are possible. A second binding site, involving a second Ca^{2+}, in each CRD has, in fact, been observed in crystals containing a high level of saccharide (Ng *et al.*, 1996). If this site were occupied, a proper model would have to include averaging of the orientation of the sugar in this site. Note that the second and third cases would preserve the axially symmetric nature of the determined order tensor but the direction of highest order will vary. A fourth possibility is that the binding geometry in the primary site can vary.

In Figure 4 we show a geometry of AMM in the binding site of MBP (1KMB.pdb) that is consistent with our determined axis of symmetry. Using this refined structure, α-methyl mannose was docked to the binding calcium of the CRD of MBP. Orientational constraints required placing AMM's experimentally derived direction of highest order along MBP's 3-fold axis of symmetry. Subsequently, rotations around this axis and translation along *x*, *y*, and *z*, while maintaining a 2.6 Å distance from the Ca^{2+} provided several possible orientations of AMM in MBP such as that depicted in Figure 4. There is a single violation of allowed van der Waals contacts (<2.5 Å). This is between HO-3 and Glu193 of MBP. The trimeric form used in our modeling has no ligand of AMM in the binding site and Glu193 coordination of the calcium compensates for the absence of sugar hydroxyl. Hence, it is quite likely that Glu193 moves on sugar binding. MBP actually has a rather open binding site, in which coordination of the three and four oxygen atoms of the sugar to Ca^{2+} are the primary interactions and considerable variation in the actual geometry of coordination may be possible.

An independent determination of MBP's order tensor would remove some of the uncertainties in the above discussion. First, this would allow determination of the relative domain orientations of MBP in solution. This can be achieved by equating the observed direction of highest order as seen from any single domain to the symmetry axis of trimeric MBP (Al-Hashimi *et al.*, 1999). This procedure, in fact, introduces a novel and simple route into determination of relative domain orientations in certain types of symmetry related homo-multimers such as MBP, a problem that so far has been

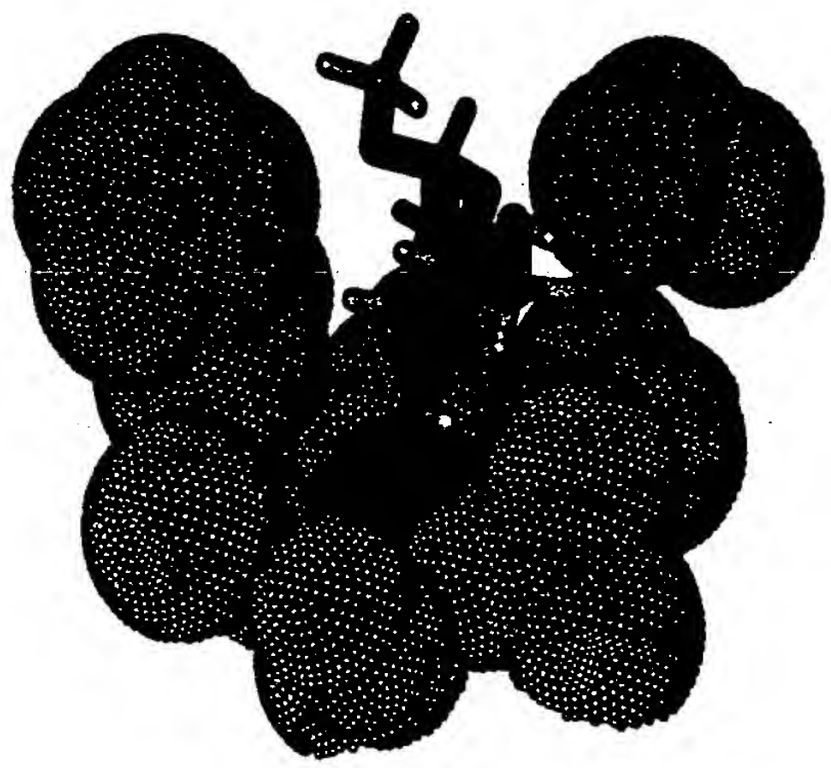


Figure 4. Proposed mode of binding of AMM in MBP. The dark blue corresponds to Ca^{2+}, and the light blue represents residues in the binding pocket of MBP.

very difficult to address using standard NMR techniques. Second, the determined direction of highest order and the corresponding order parameters will have absorbed any internal motional averaging contributions within MBP permitting direct comparison with an order tensor determined for AMM in the bound state. Additional averaging in the ligand site, either due to mobility in the bound state or to the presence of alternative binding sites, [illegible] parameter elements [illegible] of data required is [illegible] protein carbohydrate complexes where NOE-based methods suffer from lack of proximate non-exchangeable protons on the ligand and the protein. The effects of symmetry on measured residual dipolar couplings provides a powerful orientation reference for not only ligand orientation, but possibly also for the relative orientation of domains in homo-multimeric complexes.

Materials and Methods

Preparation of MBP-A

MBP-A is an expression product from the pIN-IIIompA-2 plasmid (Drickamer, 1989), and was expressed as described (Weis *et al.*, 1991). Briefly, competent JA221 cells were transformed with DNA plasmid and colonies selected from LB-amp plates for use in a starter culture in an LB-amp medium. Starter culture was grown to saturation overnight at 37 °C and used to inoculate five liters of LB-amp medium, which grew to A_{550} of 0.8. MBP-A production was then induced with IPTG and additional growth allowed for 2.5 hours. Cells were harvested by centrifugation at 4000 *g*, the pellet was resuspended in 10 mM Tris-Cl (pH 7.0) and then lysed by sonication. The insoluble pellet from 18,000 *g* centrifugation was solubulized in 6 M guanidinium chloride, 0.1 M Tris-Cl (pH 7.0), and clarified by centrifuging at 138,000 *g* for 30 minutes at 4 °C. Dialysis of the centrifuge extract with 25 mM Tris-Cl, 1.25 M NaCl, 25 mM $CaCl_2$ over 48 hours and centrifuging at 138,000 *g* provided the dialysate which was purified by affinity chromatography on a mannose-Sepharose column. Elution with 25 mM Tris-Cl, 1.25 M NaCl and 2.5 mM Na_2EDTA followed by reconstitution by dialysis against 10 mM NaCl, 1 mM Tris-Cl, 25 mM $CaCl_2$ afforded 10 mg of MBP-A from a five liter growth culture.

Preparation of AMM

AMM was synthesized from [$^{13}C_6$] glucose by minor modification of a described method (Hare *et al.*, 1993). Briefly, uniformly ^{13}C-labeled D-glucose was protected as the penta-acetate with acetic anhydride, and the crude material brominated with HBr/AcOH. Reduction with Zn/aqueous AcOH provided the desired tri-*O*-acetyl-D-glucal, which was used in a Ferrier reaction with methanol/Dowex H^+ to give (α-methyl-2,3-dideoxy-4,6-di-*O*-acetyl-D-glucopyranoside. Purification by silica gel chromatography; followed by treatment of the anhydrous material with OsO_4 in pyridine afforded the desired diacetylated α-methyl mannoside. Flash chromatography followed by deacetylation with NaOMe/MeOH provided AMM in an 20 % overall yield.

NMR spectroscopy

Two samples were prepared: (1) a 1 mM MBP-A, 1 mM AMM, 10 mM NaCl, 1 mM Tris-Cl, 25 mM $CaCl_2$ (pH 7) in 500 µL of 5 % (w/v) bicelle (DMPC/DHPC 3:1 molar ratio) solution in ^2H_2O and (2) a 2 mM AMM, 10 mM NaCl, 1 mM Tris-Cl, 25 mM $CaCl_2$ (pH 7) solution in 5 % (w/v) bicelle (DMPC/DHPC 3:1) control sample. These were prepared as described (Losonczi & Prestegard, 1998) using one bicelle stock solution. All NMR experiments described below were conducted on an Inova 600 MHz NMR spectrometer equipped with z-axis pulsed-field gradients. In all cases, bicelle alignment was monitored by 1D ^2H NMR quadrupolar splittings as a function of temperature with maximal alignment observed at 39 °C. Spectra were acquired using normal hetero-nuclear (^1H-^{13}C) single quantum coherence (HSQC) experiments modified so that ^{13}C couplings were present in the direct proton dimension. Quadrature detection in the t_1 evolution period was accomplished using gradient coherence selection. For all experiments, 1024 points per scan, 16 scans per increment, 160 t_1 increments were acquired with a direct sweep width of 3000 Hz and an indirect sweep width of 9000 Hz. Using the same acquisition parameters, another identical HSQC experiment was acquired at a temperature of 25 °C where the bicelle medium is in an isotropic solution state.

Calculations of dipolar couplings and order tensors

Dipolar couplings were calculated as the difference between the oriented couplings ($^1J_{CH} + D_{CH}$) and the isotropic couplings ($^1J_{CH}$). In each case, couplings were extracted using a Bayesian time-domain NMR parameter estimation program Xrambo, using the method described (Andrec & Prestegard, 1998). This program is available on the Internet at http://tesla.ccrc.uga.edu. Typically, a 2D HSQC data set with the desired splittings in the direct dimension was transformed and phased to yield a 1024 × 256 real matrix. Single FID slices across the width of the resonance under investigation were analyzed independently. For every slice, the resulting frequency domain data were reverse Fourier transformed to generate a 1D time domain data set as input for Xrambo. The following model was used for the data. Each component of the doublet resulting from C-H couplings was given an identical linewidth and intensity, but an independent phase to circumvent any problems resulting from the presence of phase twist anomalies or dispersive contributions. Values for shifts, linewidths, phases and intensities were estimated and entered as starting parameters which were subsequently refined by Xrambo's Metropolis Monte Carlo method. This procedure was repeated for several indirect slices of the same doublet and the resulting rmsd across various slices was used as the uncertainty in measurement. The error analysis from this procedure was used to estimate the final precision of all measured residual dipolar couplings. The measured residual dipolar couplings and associated uncertainties along with AMM input coordinates obtained from an MD simulation were then used as input to a singular value decomposition program for the determination of order tensor elements (Losonczi *et al.*, 1999). This

program is also available on the internet at http://tesla.ccrc.uga.edu.

Molecular dynamics simulation

MD simulations were performed with AMBER 4.1 (Pearlman *et al.*, 1995) employing the all-atom GLYCAM93 parameter set for oligosaccharides (Woods *et al.*, 1995) on a 12 processor SGI Origin 2000 computer. AMM was solvated in an approximately 32 Å × 32 Å × 32 Å box of 466 TIP3P water molecules and was minimized by steepest descent energy minimization using a dielectric constant of 1.0 and a cut-off value for non-bonded pair interactions of 8.0 Å. Initial atomic velocities were assigned at 5 K, and the simulation run at 300 K at 1 atm. Selected torsions were extracted with the CARNAL module of AMBER 4.1 from the trajectories to determine a closely related family of the most populated structures.

Docking studies

First approximations of AMM's position in MBP were done with MIDAS Plus (Ferrin *et al.*, 1988). Briefly, MBP with AMM modeled into the CRD were aligned with the helix parallel with the z-axis. A second AMM was matched and oriented according to the singular value decomposition analysis of the calculated dipolar couplings (a $z-10°$, $y+40°$ rotation from the crystal structure). This was then translated in x, y, and z directions while limiting rotations to the z-axis only. Solutions that maintained the three and four OH groups 2.6-2.8 Å from the Ca^{+2} responsible for binding and that minimized van der Waals interaction with the protein were deemed plausible solutions.

Acknowledgments

The authors thank Eric Sayers for help in MBP-A preparation, M. W. F. Fischer for useful discussions on order tensor analysis of MBP, and Laura Morris for help with molecular dynamics simulations. This work was supported by NIH grants GM33225 and RR05351.

References

Al-Hashimi, H. M., Bolon, P. J. & Prestegard, J. H. (1999). Molecular symmetry effects on experimentally determined order tensors for exchanging ligands. *J. Magn. Reson.* In the press.

Andrec, M. & Prestegard, J. H. (1998). Metropolis Monte Carlo implementation of Bayesian time-domain parameter estimation: application to coupling constant estimation from antiphase multiplets. *J. Magn. Reson.* **130**, 217-232.

Bolon, P. J. & Prestegard, J. H. (1998). COSY cross-peaks from ^{1}H-^{1}H dipolar couplings in NMR spectra of field oriented oligosaccharides. *J. Am. Chem. Soc.* **120**, 9366-9367.

Bothner-By, A. A. (1995). Magnetic field induced alignment of molecules. In *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D. M. & Harris, R. K., eds), pp. 2932-2938, Wiley, Chichester.

Bugayevskiy, L. M. & Snyder, J. P. (1995). *Map Projections: A Reference Manual*, Taylor & Francis, London.

Chang, C. Y., Sastry, K. N., Gillies, S. D., Ezekowitz, R. A. B. & Sheriff, S. (1994). Crystallization and preliminary X-ray analysis of a trimeric form of human mannose binding protein. *J. Mol. Biol.* **241**, 125-127.

Clore, G. M., Gronenborn, A. M. & Bax, A. (1998a). A robust method for determining the magnitude of the fully asymetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133**, 216-221.

Clore, G. M., Gronenborn, A. M. & Tjandra, N. (1998b). Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159-162.

Clore, G. M., Starich, M. R. & Gronenborn, A. M. (1998c). Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.* **120**, 10571-10572.

Drickamer, K. (1989). Demonstration of carbohydrate-recognition activity in diverse proteins which share a common primary structure motif. *Biochem. Soc. Trans.* **17**, 13-15.

Dwek, R. A. (1996). Glycobiology: toward understanding the function of sugars. *Chem. Rev.* **96**, 683-720.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display System. *J. Mol. Graph.* **6**, 13-27, 36-37.

Fischer, M. W. F., Losonczi, J. A., Weaver, J. L. & Prestegard, J. H. (1999). Domain orientation and dynamics in multi-domain proteins from residual dipolar couplings. *Biochemistry*, **38**, 9013-9012.

Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065-1074.

Hare, B. J., Sanders, C. R., McIntyre, S. E. & Prestegard, J. H. (1993). Synthesis and characterization of a ^{13}C-labeled alpha-mannosyl glycolipid analog from [^{13}C]glucose. *Chem. Phys. Lipids*, **66**, 155-158.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T. & Yamashina, I. (1987). Serum lectin with known structure activates complement through the classical pathway. *J. Biol. Chem.* **262**, 7451-7454.

Iobst, S. T., Wormald, M. R., Weis, W. I., Dwek, R. A. & Drickamer, K. (1994). Binding of sugar ligands to Ca^{2+}-dependent animal lectins. I. Analysis of mannose binding by site-directed mutagenesis and NMR. *J. Biol. Chem.* **269**, 15505-15511.

Kiddle, G. R. & Homans, S. W. (1998). Residual dipolar couplings as new conformational restraints in isotropically 13C-enriched oligosaccharides. *FEBS Letters*, **436**, 128-130.

Kolatkar, A. R., L., A., Isecke, R., Brossmer, R., Drickamer, K. & Weis, W. I. (1998). Mechanism of *N*-acetylgalactosamine binding to a C-type animal lectin carbohydrate-recognition domain. *J. Biol. Chem.* **273**, 19502-19508.

Kung, H. C., Wang, K. Y., Goljer, I. & Bolton, P. H. (1995). Magnetic alignment of duplex and quadruplex DNAs. *J. Magn. Reson. ser. B*, **109**, 323-325.

Lasky, L. A. (1992). Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science*, **258**, 964-969.

Losonczi, J. A. & Prestegard, J. H. (1998a). Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J. Biol. NMR*, **12**, 447-451.

Losonczi, J. A. & Prestegard, J. H. (1998b). Nuclear magnetic resonance characterization of the myristoy-

lated, N-terminal fragment of ADP-ribosylation factor 1 in a magnetically oriented membrane array. Biochemistry, 37, 706-716.

Losonczi, J. A., Andrec, M., Fischer, M. W. F. & Prestegard, J. H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. J. Magn. Reson. 138, 334-342.

[illegible]

[illegible] AMBER 4.1, University of California, San Francisco.

Prestegard, J. H. (1998). New techniques in structural NMR — anisotropic interactions. Nature Struct. Biol. 5, 517-522.

Prestegard, J. H., Tolman, J. R., Al-Hashimi, H. M. & Andrec, M. (1998). Protein structure and dynamics from field induced residual dipolar couplings. In *Modern Techniques in Protein NMR* (Krishna, N. R. & Berliner, L. J., eds), Plenum, New York.

Rundlof, T., Landersjo, C., Lycknert, K., Maliniak, A. & Widmalm, G. (1998). NMR investigation of oligosaccharide conformation using dipolar couplings in an aqueous dilute liquid crystalline medium. *Magn. Res. Chem.* 36, 773-776.

Sanders, C. R., Hare, B. J., Howard, K. P. & Prestegard, J. H. (1994). Magnetically-oriented phospholipid micelles as a tool for the study of membrane-associated molecules. Prog. NMR Spectrosc. 26, 421-444.

Saupe, A. (1968). Recent results in the field of liquid crystals. Angew. Chem. Int. Ed. Engl. 7, 97-112.

Sayers, E. W. [illegible]

[illegible] information for structure determination in solution. Proc. Natl Acad. Sci. USA, 92, 9279-9283.

Weis, W. I. & Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. Structure, 2, 1227-1240.

Weis, W. I., Crichlow, G. V., Murthy, H. M. K., Hendrickson, W. A. & Drickamer, K. (1991). Physical characterization and crystallization of the carbohydrate-recognition domain of a mannose-binding protein from rat. *J. Biol. Chem.* 266, 20678-20686.

Woods, R. J., Dwek, R. A., Edge, C. J. & Fraser-Reid, B. (1995). Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem.* 99, 3832-3846.

Edited by P. E. Wright

(Received 18 May 1999; received in revised form 17 August 1999; accepted 18 August 1999)

APPENDIX 3

RELATED PROCEEDINGS APPENDIX

None